(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) EP 0 977 175 B1

(12) EUROPEAN PATENT SPECIFICATION

(45) Date of publication and mention of the grant of the patent:
09.11.2005 Bulletin 2005/45

(51) Int Cl.7: **G10L 15/18**

(21) Application number: **99305887.4**

(22) Date of filing: **26.07.1999**

(54) Method and apparatus for recognizing speech using a knowledge base

Verfahren und Vorrichtung zur Spracherkennung unter Verwendung einer Wissensbasis

Procédé et dispositif de reconnaissance de la parole utilisant une base de connaissance

(84) Designated Contracting States:
DE FR GB IT NL

(30) Priority: **27.07.1998 JP 21097998**
15.07.1999 JP 20152899

(43) Date of publication of application:
02.02.2000 Bulletin 2000/05

(73) Proprietor: **CANON KABUSHIKI KAISHA**
Tokyo (JP)

(72) Inventors:
- **Yamamoto, Naoko, c/o Canon Kabushiki Kaisha**
  Tokyo (JP)
- **Suda, Aruna Rohra, c/o Canon Kabushiki Kaisha**
  Tokyo (JP)
- **Wakai, Masanori, c/o Canon Kabushiki Kaisha**
  Tokyo (JP)

(74) Representative:
Beresford, Keith Denis Lewis et al
BERESFORD & Co.
16 High Holborn
London WC1V 6BX (GB)

(56) References cited:
EP-A- 0 344 017 **US-A- 4 937 870**

- **KAWAHARA T ET AL: "HEURISTIC SEARCH INTEGRATING SYNTACTIC, SEMANTIC AND DIALOG-LEVEL CONSTRAINTS" PROCEEDINGS OF THE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, ANDSIGNAL PROCESSING. (ICASSP),US,NEW YORK, IEEE, vol. CONF. 19, 1994, pages II-25-II-28, XP000528443 ISBN: 0-7803-1776-9**
- **GLASS J ET AL: "MULTILINGUAL SPOKEN-LANGUAGE UNDERSTANDING IN THE MIT VOYAGER SYSTEM" SPEECH COMMUNICATION,NL,ELSEVIER SCIENCE PUBLISHERS, AMSTERDAM, vol. 17, no. 1/02, 1 August 1995 (1995-08-01), pages 1-18, XP000642167 ISSN: 0167-6393**
- **HITOSHI IIDA ET AL: "PREDICTING THE NEXT UTTERANCE LINGUISTIC EXPRESSIONS USING CONTEXTUAL INFORMATION" IEICE TRANSACTIONS ON INFORMATION AND SYSTEMS,JP,INSTITUTE OF ELECTRONICS INFORMATION AND COMM. ENG. TOKYO, vol. E76-D, no. 1, 1 January 1993 (1993-01-01), pages 62-73, XP000354887 ISSN: 0916-8532**
- **PATENT ABSTRACTS OF JAPAN vol. 017, no. 419 (P-1585), 4 August 1993 (1993-08-04) & JP 05 080793 A (KOKUSAI DENSHIN DENWA CO LTD ;OTHERS: 01), 2 April 1993 (1993-04-02)**
- **PATENT ABSTRACTS OF JAPAN vol. 1998, no. 07, 31 March 1998 (1998-03-31) & JP 01 088600 A (TOSHIBA CORP), 3 April 1989 (1989-04-03)**
- **SETLUR A R ET AL: "Correcting recognition errors via discriminative utterance verification" PROCEEDINGS ICSLP 96. FOURTH INTERNATIONAL CONFERENCE ON SPOKEN LANGUAGE PROCESSING (CAT. NO.96TH8206), PROCEEDING OF FOURTH INTERNATIONAL CONFERENCE ON SPOKEN LANGUAGE PROCESSING. ICSLP '96, PHILADELPHIA, PA, USA, 3-6 OCT. 1996, pages 602-605 vol.2, XP002139609 1996, New York, NY, USA, IEEE, USA ISBN: 0-7803-3555-4**

Note: Within nine months from the publication of the mention of the grant of the European patent, any person may give notice to the European Patent Office of opposition to the European patent granted. Notice of opposition shall be filed in a written reasoned statement. It shall not be deemed to have been filed until the opposition fee has been paid. (Art. 99(1) European Patent Convention).

EP 0 977 175 B1

Description

BACKGROUND OF THE INVENTION

Field of the Invention

[0001] The present invention broadly relates to speech recognition and, more particularly, to a method and an apparatus for recognizing speech information based on prediction concerning an object to be recognized. The invention also relates to a storage medium for storing a program implementing the above method.

Description of the Related Art

[0002] Speech recognition is primarily divided into two types of methods, i.e., a word speech-recognition method and a clause speech-recognition method. According to the word speech-recognition method, an input speech waveform is analyzed, and features are extracted from the waveform to produce a feature time series. Then, the similarity of the features in relation to the word dictionary represented by the feature time series which has been similarly obtained is calculated, and the calculated word is output as a recognition result. In the clause speech-recognition method, input speech is converted into phoneme strings, which are substituted by word strings. The word strings are then parsed, and are converted into character strings. Logic analyses and semantic analyses are then made on the character strings, so that a sentence is produced and output. Further research is being conducted on a method of providing word class information for homonyms, and a method of converting input speech into compound nouns or into a single clause. It is however very difficult to implement such methods.

[0003] In most cases, during conversation, humans recognize the speaker's voice by understanding it as one meaning. While the speaker is speaking, the listener supports his/her understanding by predicting the content of the speech to some degree according to the previous topic and common sense. Consequently, even if the speaker wrongly selects or pronounces some words, the listener understand him/her without any problem. Even if there are many homonyms in a conversation, the listener can determine which word the speaker means.

[0004] In contrast, conventional speech recognition systems perform speech recognition according to pattern matching. More specifically, a dictionary provided for a system is searched for possible words which match a certain portion of an input speech waveform, and the searched words are output. Among the output words, the optimal word is selected. With this arrangement, if speech recognition fails while it is being conducted, the subsequent processing is spoilt.

[0005] Additionally, in most conventional speech recognition systems, it is assumed that input speech to be recognized satisfies the syntax of a certain language. Thus, various determinations are made in a speech recognition module, and the determination result is transferred to another process (another module). More specifically, in a speech recognition module, speech information is uniquely determined as a system command by being filtered (parsed). Not only processing for grammatically correct speech, but also processing for unnecessary words, such as exclamations and restated words, and for non-grammatical speech, such as anastrophy (inversion) and particle dropping is handled by language processing (verifying such words against a word database or a grammar database).

[0006] However, since parsing is performed in order to analyze the structure of syntax, elements other than syntax information are rejected. Even if a word is determined to be a significant word after parsing, general knowledge or knowledge of a specific field is not considered.

[0007] An example of conventional speech recognition systems is shown in Fig. 42. Since the flow of processing executed on input speech is unidirectional, the system processing Continues to proceed in the same direction even if the processing result of a speech recognition module is incorrect. For example, an input that is determined to be syntactically correct but cannot be processed by the entire system upon performing speech recognition is disadvantageously received, and returns as an error. That is, a speech recognition unit and the whole system separately perform processing without operating together, thereby failing to implement complicated processing. As a consequence, the performance of the entire system is seriously influenced by the result of speech recognition.

[0008] JP-A-05080793 describes an interactive understanding device which includes a speech recognition device and a word predictor which operates to constrain a vocabulary used by the speech recognition device in order to improve recognition efficiency.

SUMMARY OF THE INVENTION

[0009] According to one aspect, the invention provides a speech information processing apparatus comprising:

a context base for storing prediction information representing a category of words to be subsequently recognised;

recognition means for recognising a word based on the category of words represented by the prediction information stored in said context base;

a knowledge base for storing knowledge concerning a category of speech information;

prediction means for predicting the category of words to be subsequently recognised based on at least one previously recognised word by referring to the knowledge stored in said knowledge base;

and

updating means for updating the prediction information stored in said context base based on the category of words obtained by said prediction means;

characterised in that:

the apparatus further comprises determination means for determining whether or not input sound information is language information by referring to the knowledge stored in said knowledge base; and in that

said recognition means is arranged to perform the recognition on the input sound information when said determination means determines that the input sound information is language information.

[0010] According to another aspect, the present invention provides a speech information processing method comprising:

a recognition step of recognising a word based on a category of words to be subsequently recognised which is represented by prediction information stored in a context base;

a prediction step of predicting the category of words to be subsequently recognised based on at least one previously recognised word by referring to knowledge concerning a category of speech information stored in a knowledge base concerning a category of speech information; and

an updating step of updating the prediction information stored in said context base based on the category of words obtained in said prediction step;

characterised by a determination step of determining whether or not input sound information is language information by referring to the knowledge stored in said knowledge base; and in that said recognition step performs the recognition on the input sound information when the input sound information is determined to be language information in said determination step.

[0011] Other objects and advantages besides those discussed above shall be apparent to those skilled in the art from the description of a preferred embodiment of the invention which follows. In the description, reference is made to accompanying drawings, which form a part thereof, and which illustrate an example of the invention. Such example, however, is not exhaustive of the various embodiments of the invention.

BRIEF DESCRIPTION OF THE DRAWINGS

[0012]

Fig. 1 illustrates the configuration of the hardware of a natural-language processing apparatus according to an embodiment of the present invention;
Fig. 2 illustrates a system architecture;
Fig. 3 illustrates an implementation mode in speech recognition;
Fig. 4 illustrates the outline of understanding speech issued by humans;
Fig. 5 illustrates input processing;
Fig. 6 illustrates the configuration of a system;
Fig. 7 illustrates a schematic flow of the system processing;
Fig. 8 is a flow chart schematically illustrating the processing performed by the entire apparatus;
Fig. 9 is a flow chart illustrating analyzing procedure of the process result;
Fig. 10 is a flow chart illustrating the recognition processing reflecting prediction information;
Fig. 11 is a flow chart illustrating the flow of speech recognition processing;
Fig. 12 is a flow chart illustrating the flow of determining the type of input sound;
Fig. 13 illustrates two prediction techniques;
Figs. 14 and 15 illustrate the classification of the categories of words;
Fig. 16 is a flow chart illustrating the procedure of setting the initial prediction;
Fig. 17 is a flow chart illustrating the word-recognition processing;
Fig. 18 is a flow chart illustrating the syllable-recognition processing;
Fig. 19 is a flow chart illustrating matching processing between the syllable recognition result and the word provided for the system;
Figs. 20 and 21 are flow charts illustrating the processing of determining the similarity of syllables;
Fig. 22 is a flow chart illustrating the processing of calculating the similarity of the corresponding word by utilizing the similarity of syllables and the recognition time;
Fig. 23 is a flow chart illustrating the indication processing;
Fig. 24 illustrates the parameter-setting/result-indication screen;
Fig. 25 illustrates an example of a syllable dictionary;
Fig. 26 illustrates an example of a word dictionary;
Fig. 27 illustrates the state transition of the context for a prediction of a subsequent input;
Fig. 28 is a flow chart illustrating the processing of generating a response to the user;
Fig. 29 illustrates an example of a language dictionary;
Fig. 30 illustrates an example of a concept dictionary;
Fig. 31 illustrates an example of rules;
Fig. 32 illustrates an example of a word dictionary;
Fig. 33 illustrates an example of syllable recognition results;
Fig. 34 illustrates a dynamic programming (DP) matching algorithm;
Fig. 35 illustrates an example of a word dictionary;
Fig. 36 illustrates an example of syllable recognition results;

Fig. 37 is a flow chart illustrating the processing of determining a recognition result and of determining whether the recognition result is to be accepted;
Fig. 38 is a flow chart illustrating the recognition-result analyzing processing;
Fig. 39 is a flow chart illustrating a concept analysis and the processing of determining an analysis result;
Fig. 40 is a flow chart illustrating the result correction processing;
Fig. 41 is a flow chart illustrating the processing of re-determining the previous recognition result; and
Fig. 42 illustrates conventional input processing.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

[0013] Preferred embodiments of the present invention are described in detail below with reference to the accompanying drawings.

First Embodiment

[0014] A detailed description is given below of a first embodiment of the present invention with reference to the drawings.

[0015] A discussion is first given of the construction of the hardware for use in a natural-language processing apparatus according to the first embodiment of the present invention. Referring to the block diagram illustrating the construction of the hardware shown in Fig. 1, an input unit 101 inputs information through the use of a natural language. It is not essential that the input information be a grammatically complete sentence as long as it has a regular structure.

[0016] The input unit 101 is not limited to a speech recognition system for inputting and recognizing speech, and may be a keyboard for inputting characters through keys, a character recognition reader for optically reading characters from a document and recognizing them, an online/offline handwritten-character recognition reader, or a receiving unit for receiving information from another system, for example, from a character recognition system. Alternatively, two of the above input units may be combined and selectively utilized as the input unit 101.

[0017] A CPU 102 conducts calculations and Boolean operations for various processing, and controls the individual elements connected to a bus 106. An output unit 103 outputs analyzed data information, and may be a speech synthesizing unit for synthesizing speech from character information and outputting it, a display unit, such as a cathode ray tube (CRT) or a liquid crystal display unit, a printer for printing characters on a document, or a transmitter for transmitting information to another unit, such as a database. The output from the output unit 103 may be input into another output unit within the same apparatus, for example into a concept analyzing unit. Alternatively, two of the above-described units may be combined and selectively utilized as the output unit 103.

[0018] A program memory 104 stores a program including the processing procedure controlled by the CPU 102, which will be described below with reference to a flow chart. The program memory 104 may be a read only memory (ROM) or a random access memory (RAM) into which the program is loaded from an external storage device.

[0019] A data memory 105 stores not only data generated by various processing, but also a knowledge base, which will be discussed below. The data memory 105 may be a RAM, but knowledge included in the knowledge base is loaded into the data memory 105 from a non-volatile external storage medium before processing is executed, or is checked every time the need arises. The bus 106 is used for transmitting address signals which give an instruction to the individual elements controlled by the CPU 102 and for transferring data to be exchanged between the individual units.

[0020] Fig. 2 is a block diagram illustrating the basic configuration of an information processing apparatus according to the first embodiment. The information processing apparatus performs processing by using knowledge of the knowledge base. Fig. 2 illustrates the flow of the processing executed by using this knowledge.

[0021] The information processing apparatus includes an input processing unit 201 for executing processing on the individual input signals so as to obtain input information. A context-construction/goal-inference unit 202 conducts concept-analyses on the content of natural-language information input from the input unit 101 by utilizing the knowledge of a knowledge base 208, thereby understanding the meaning.

[0022] A planning unit 203 makes planning by using a context base 207 and the knowledge of the knowledge base 208 in order to achieve a goal inferred by the context-construction/goal-inference unit 202.

[0023] An execution unit 204 requests, based on the processing result of the planning unit 203, a main application unit 205 to execute processing by using an application, a database, or a printer connected to a system. The main application unit 205 then executes processing by using an application, database, or a printer connected to the system.

[0024] A response determining unit 206 receives the processing result of the execution unit 204 and determines a response to be output to a user. In this embodiment, the response determining unit 206 analyzes the output by employing the context base 207 and the knowledge of the knowledge base 208 and generates a response if required, and finally selects a method for outputting the response.

[0025] The context base 207 provides the knowledge required for the context-construction/goal-inference unit 202, the planning unit 203, and the response determin-

ing unit 206, and also stores new knowledge generated while the above units are executing processing.

[0026] The knowledge base 208 provides the knowledge required for the context-construction/goal-inference unit 202, the planning unit 203, and the response determining unit 205, and also stores new knowledge produced while the above units are executing processing.

[0027] Fig. 3 illustrates the flow of the processing performed by the information processing apparatus of the first embodiment. An input recognition unit 301, which corresponds to the input processing unit 201 shown in Fig. 2, recognizes the input information.

[0028] A concept analyzing unit 302, which corresponds to the context-construction/goal-inference unit 202, the planning unit 203, and the execution unit 204, analyzes the meaning of the input information by utilizing knowledge-base/context-base 306 contained in the system according to the recognition result of the input recognition unit 301. Upon analyses, the concept analyzing unit 302 predicts information to be subsequently input or requests a main application unit 303 to execute processing.

[0029] The main application unit 303, which corresponds to the main application unit 205, executes processing requested by the concept analyzing unit 302 and transfers the execution result to a response generating unit 304.

[0030] The response generating unit 304, which performs processing on the result of the response determining unit 206, analyzes the execution result of the main application unit 303 and generates a response to be output to the user, and also selects the optimal output method.

[0031] The response generating unit 304 requests an output synthesizing unit 305 to output the response. The output synthesizing unit 305 outputs the response generated by the response generating unit 304 according to the selected method. The knowledge-base/context-base 306 of the system is used for performing processing by the response generating unit 304 and the output synthesizing unit 305.

[0032] By applying the construction of the information processing apparatus of the first embodiment to speech recognition, the advantages of human speech recognition processing are implemented in this apparatus. An example of a mechanism of recognizing speech issued by humans is given below. In this example, it is assumed that input speech "Send mail to May" is processed.

[0033] Fig. 4 illustrates the outline of understanding the speech "Send mail to May" by humans. In most cases, humans recognize the speech by understanding it as one meaning rather than by sequentially selecting possible words similar to a certain portion of the input speech waveform, as performed in current speech recognition systems. This is because humans recognize and understand speech not only by speech information, but also by predicting the context used before and after the speech and common sense to some extent.

[0034] In order to implement the humans' recognition operation in a system, predictions may be made on the input information in advance. More specifically, when "Send mail to May!" is input as speech, the following predictions are made. Upon recognizing the word "send", the subsequent object is predicted by using the language knowledge, and a word "mail" is further predicted by using the domain knowledge.

[0035] Generally, in speech recognition, possible words "male", "mail", "may", and "mate" may be searched for. Among these words, "mail" may be predicted as a personal name from the language knowledge, which is unlikely to be contained in an ordinary dictionary. However, since personal names are not likely to come immediately after "send", "mail" is rejected. Also, a stereotyped phrase "send to" is predicted from the language knowledge, and "mate" is not likely to be selected. Further, "to", which is a homonym of "to", is not predicted from the knowledge base. Finally, it is predicted from the concept knowledge that XXX in "send to XXX" may be an object, and from the domain knowledge that the destination of "send to" may be a human (personal name). It is thus considered that "May" be predicted from an address book or a biographical dictionary.

[0036] A comparison is then made between the speech recognition processing shown in Fig. 4 and the processing performed by the conventional speech recognition system shown in Fig. 42.

[0037] According to conventional input processing methods, various determinations are generally made in a recognition module, and the result is transferred to another module. The input information is recognized by executing the recognition processing and is shaped into a form receivable by an application. The flow of the processing is unidirectional, and the individual units separately perform the processing rather than operating together.

[0038] Particularly for the processing of input speech, the following method is usually employed, as illustrated in Fig. 42. The result obtained by recognizing the speech in a speech recognition unit 4201 is uniquely determined as a system command by being filtered (parsed) in a speech processing unit 4202. Not only processing for grammatically correct speech, but also processing for unnecessary words, such as exclamations and restated words, and for non-grammatical speech, such as anastrophy (inversion) and particle dropping is handled by language processing (verifying such words against a word database or a grammar database) in the speech processing unit 4202. Parsing is performed in order to analyze the structure of syntax, and accordingly, elements other than syntax information, which may also be useful, are rejected. Additionally, the flow of the processing is unidirectional, and even if the processing result of the speech recognition module is incorrect, the system completes the processing performed by a speech input unit 4203 and proceeds to a subsequent stage, i.e., an

application unit 4207. Processing is similarly performed in a keyboard input and an image input unit 4206 (an optical character reader (OCR) 4204 and an image processing unit 4205).

[0039] According to the aforementioned method, even an input which cannot be processed by the entire system is accepted, and is returned as an error from the application unit 4207. That is, the processing on speech and images performed by the input units 4203 and 4206 is not operated together with the processing of the entire system, thereby implementing merely a simple operation. As a result, the performance of the entire system is seriously influenced by the result of speech recognition.

[0040] In contrast, an input processing method according to the first embodiment is shown in Fig. 5. Input information, if it indicates sound, is recognized in a speech recognition unit 501, and if it represents an image, is recognized in an OCR 502. The result is then analyzed in an analyzing unit 503 based on common sense and knowledge, and a subsequent input is further predicted or the result of analyses is transferred to an application unit 504 of the system.

[0041] Particularly for processing input speech, the speech recognition is ideally performed by comprehensively using speech information and other knowledge rather than solely performing speech recognition as in a conventional manner. The result obtained by processing speech is stored as the knowledge of the whole system, and is used together with the other knowledge included in the system, thereby making it possible to recognize the meaning of the speech rather than the structure of the speech. That is, according to the flow of the processing indicated by the arrows 505 and 506 shown in Fig. 5, the results of analyses are fed back to the speech recognition unit 501 and the OCR 502 so that the recognition result and the analyses results are operated together, thereby improving the performance of the input processing. According to the flow of the processing indicated by the arrows 507 and 508 illustrated in Fig. 5, the analysis result and the processing of the application unit 504 are operated together, thereby enhancing the performance of execution processing. As a consequence, the performance of the entire system can be improved.

[0042] In this embodiment, the input processing illustrated in Fig. 5 and the system architecture shown in Fig. 2 are implemented. As a result, processing similar to the speech recognition processing performed by humans can be achieved. More specifically, Fig. 6 is an overall diagram illustrating a speech recognition system constructed in accordance with an implementing method, such as that shown in Fig. 3.

[0043] When speech is input, speech recognition is conducted by a speech recognition unit 601 according to the previously made predictions and information contained in a knowledge-base/context-base 606. For the processing of speech information, the knowledge-base/context-base 606 includes not only common knowledge effective for the processing regardless of the type of knowledge, but also knowledge concerning speech information. The concept of the recognition result is analyzed by a concept analyzing unit 602 by utilizing the common sense and the knowledge of the system contained in the knowledge-base/context-base 606, thereby analyzing the meaning of the recognition result.

[0044] A main application unit 603 predicts a subsequent speech input or performs processing according to the purpose. Upon executing the processing by the main application unit 603, a response to the user may be required, in which case, a response is generated in a response generating unit 604. If it is determined that a response is most suitably given to the user in speech, a response is converted into speech in a speech synthesizing unit 605 and is output. The knowledge-base/context-base 606 of the system is also used for the above processing.

[0045] There are primarily two techniques for predicting speech, and the details are shown in Fig. 13. Predictions may be made at two stages. When speech is recognized, a subsequent input signal may be predicted. Alternatively, when the output result of a speech recognition engine is used for internal processing, a result to be subsequently input may be predicted.

[0046] According to the first technique, a word to be subsequently input is predicted from previously input words and common sense by utilizing a knowledge base. Speech (phonemes or syllables) to be subsequently input is further predicted from the predicted word, and is utilized for enhancing the speech recognition rate. According to the second technique, a word to be subsequently input is also predicted from previously input speech and common sense by utilizing the knowledge base, and is used for smoothly performing subsequent processing.

[0047] For example, if a domain represents an apparatus for transmitting documents or mail, the state transition of the context illustrated in Fig. 27 is predicted. An initial prediction is made in the apparatus as follows. First, a prediction is made from the general knowledge, such as "a user may take some action in order to operate the apparatus", that a verb is likely to be input. Then, verbs that may be accepted by this apparatus are categorized as Act, and the apparatus waits for an input of speech by predicting that a verb belonging to the category Act is to be input.

[0048] After recognizing the input of a verb belonging to the category Act, the state transition of the prediction category occurs. That is, a prediction is then made on speech belonging to a category Object. The classification of the categories, such as Act and Object, is shown in, for example, Figs. 14 and 15. For example, the category Object handled in this apparatus includes mail, document, etc.

[0049] Fig. 8 is a flow chart schematically illustrating the processing performed by the entire apparatus. A

subsequent object to be recognized is predicted based on a knowledge base storing information concerning knowledge.

[0050] In step S800, the system is started. Then, in step S801, an initial prediction is set. Fig. 16 is a flow chart illustrating the procedure of setting the initial prediction. In setting the initial prediction, since information to be recognized has not yet been input, a subsequent operation is predicted based on the previous operation, and input information is predicted based on the predicted operation.

[0051] In step S1601, the previous operation is obtained by referring to the previous processing state of the system or the content of the user's request. If it is found in step S1602 that the previous operation cannot be obtained due to the absence of a previous operation, which state is encountered immediately after the system has just started, the flow proceeds to step S1608 in which an initial prediction is set in the apparatus. In this flow chart, it is determined in step S1608 that the user must request the apparatus to take some Action, and verbs are activated as a context to be recognized by the speech recognition system.

[0052] On the other hand, if it is determined in step S1602 that the previous operation has been successfully obtained, the flow proceeds to step S1603 in which the operations related to the previous operation are checked by referring to the general knowledge or the domain knowledge related to the apparatus contained in the knowledge base. Then, in step S1604, among the operations checked in step S1603, the operation which is most likely to be performed is predicted. A determination is then made in step S1605 whether the subsequent operation has been successfully predicted. If the outcome of step S1605 is yes, the flow proceeds to step S1606. In step S1606, information related to the predicted operation is acquired from the knowledge base, and in step S1607, information to be input is predicted based on the information acquired in step S1606.

[0053] For example, if the previous operation is "print three copies of a document", the operations related to the "printing operation", such as "check print status" and "print another copy", are checked in step S1603 from the knowledge base. Then, in step S1604, it can be predicted from the previous operation "print three copies" that "check print status" is more likely to be input than "print another copy". In step S1606, by referring to the domain knowledge from the operation "check print status", the related information, such as a keyword, for example, "printer status" can be obtained. By using the obtained information, subsequent input information is predicted in step S1607.

[0054] If the previous operation is "delete all the jobs", the subsequent operation cannot be obtained in step S1603 and S1604, and a determination of step S1605 becomes no. Then, an initial prediction is set in step S1608.

[0055] Referring back to Fig. 8, input information is acquired in step S802, and it is determined in step S803 whether the input information is valid. If it is valid, the process proceeds to step S804 in which the type of input information is determined. The conceivable types of information may be speech, characters, and images. In step S805, the input information is then recognized based on the prediction according to the type of information determined in step S804. More specifically, in step S805, the input information is recognized in the following manner. It is first checked whether the information is language information or non-language information, and if it is language information, the unit of information, such as a syllable or a phoneme, is determined.

[0056] The recognized result in step S805 is used for predicting subsequent information to be recognized while analyzing the process result in step S806. Fig. 9 is a flow chart illustrating the analyzing procedure of the process result. It is considered that a prediction (step S904) is made not only from the recognition result, but also from an analysis of the recognition result (step S901), a determination of the correctness of the result (step S902), and a correction of the result (step S903). It is not essential that all the processing in steps S901 through S903 be performed. Alternatively, only an analysis of the recognition result or only a correction of the result may be performed. Alternatively, a combination of a determination of the correctness of the result and a correction of the result, or a combination of an analysis of the recognition result and a determination of the correctness of the result may be performed. In step S905, according to the prediction generated in step S904, the prediction is updated for subsequent information to be recognized.

[0057] In the whole block of analyzing the recognition result in step S806, processing is executed by referring to various types of knowledge. For example, in utilizing the language knowledge, a general dictionary may be used. Thus, even if "flint" is recognized as an input signal waveform, "print" can be determined upon performing the analyzing processing, since "flint" cannot be found in the general dictionary.

[0058] In utilizing the domain knowledge, assuming that the domain is an application provided with a mail sending function, it can be predicted that "mail" is more likely to be input than "male". Also, by utilizing the common sense (general knowledge), if, for example, printing was performed as the previous command, it can be predicted that the subsequent operation may be to indicate the printer status.

[0059] After advancing the process by using the knowledge, it is determined in step S807 whether the system should execute the processing. For example, upon completing the recognition of an input sentence, if the recognized sentence indicates an instruction to the system to execute the processing, it is determined in step S807 that the corresponding processing should be executed. Then, in step S808, the processing is executed, and in step S809, a response is provided to the user

if necessary. If it is determined in step S807 that another input is to be made, the flow returns to step S802 since the prediction for the subsequent input has already been updated. In step S802, the subsequent input is obtained. Upon completing a series of processing, a determination is made in step S810 as to whether the system is to be closed or to be continued. If the system continues to be used, the process returns to step S801 in which an initial prediction is set based on the completed operation. If the system is to be closed, it is shut down in step S811.

[0060] It is now considered in what procedure input speech is actually processed in the speech recognition system of this first embodiment according to the above-described processing illustrated in Fig. 8. The schematic flow of the process of the system in response to user's speech is shown in Fig. 7.

[0061] It is first predicted in this type of system that the user may take some action, and the system waits for a verb to be input into the speech recognition unit, since action is usually started with a verb. For example, when "send" is input, it is predicted that an object of "send" is to be input, and the system thus waits for an object to be input.

[0062] In this manner, the system understands the words while predicting a word to be subsequently input. If it is determined by conducting concept analyses that the execution is possible, the corresponding processing is actually executed. When it is determined upon the execution result that a response should be returned to the user, a suitable response is generated, and the corresponding sentence is created. The sentence is then output to the user according to the optimal method. If the optimal method is to synthesize speech, the created sentence is converted into speech, which is then output to the user. For example, if mail has been successfully sent to May, a response "I sent mail to may@xxx, Successfully!", is returned.

[0063] The processing procedure of the information processing apparatus of this embodiment is discussed below through a specific example.

[0064] In this example, input speech "Send mail to May" is recognized by setting a prediction, and the corresponding processing is appropriately performed by conducting concept analyses. It is now assumed that the domain of the apparatus of this embodiment is to send mail or documents by using electronic mail.

[0065] The system is started in step S800, and an initial prediction is then set in step S801 for waiting for information to be input from the user. It is determined in step S801 that the user must request the apparatus to take some action, and then, verbs are activated as a context to be recognized by the speech recognition unit. In step S802, the information input by the user is acquired, in which case, the speech issued by the user is recognized, and the speech information is obtained.

[0066] A determination is then made in step S803 of whether the speech information is valid in relation to the reference, such as the sound level. If the input is invalid, the system waits for a valid input. If the input is valid, the type of input, in this case, speech, is determined in step S804.

[0067] Then, in the recognition processing of the input information in step S805, speech recognition processing shown in Fig. 10 is specifically performed. In step S1001, speech recognition is first conducted. The details are given below with reference to the flow chart of Fig. 11. Upon inputting speech, sound information is first processed in step S1101. Then, it is determined in step S1108 whether the sound information is language information or non-language information. This determination may be made by using a language-information database and a non-language-information database of the knowledge base or by checking the frequency of the sound information.

[0068] A specific example of this determination in step S1108 is shown in Fig. 12. In step S1201, a reference frequency range, which is set in the apparatus, for determining whether input sound has been issued by a human is obtained. If it is determined in step S1202 based on the reference frequency range obtained in step S1201 that the input sound has been issued by a human, the process proceeds to step S1208. In step S1208, even if the input sound is included in the frequency range of human speech, information concerning elements other than general language speech, for example, a laughing voice and redundant words, i.e., information concerning "specific sound", which is registered as the knowledge of the apparatus, is acquired.

[0069] Thereafter, a determination is made in step S1209 as to whether the current input is specific sound. If the outcome of step S1209 is yes, the type of input is determined in step S1210, and a flag is set in step S1211 to indicate that the input sound is non-language information. If it is determined in step S1209 that the current input is not specific sound a flag is set in step S1212 to indicate that the input speech is language information.

[0070] On the other hand, if it is determined in step S1202 that the input sound is outside the frequency range of human speech, a flag is set in step S1203 to indicate that the input sound is non-language information. In step S1204, information concerning the sound which should be specifically processed among the non-language information is acquired. It is then determined in step S1205 whether the input sound among the non-language information should be specifically processed. If the result of step S1205 is yes, the type of sound is obtained and set in step S1206. In contrast, if it is found in step S1205 that the input sound does not have to be specifically processed, the type of input sound is set to be noise in step S1207.

[0071] According to the determination process as described above, the type of speech is determined in step S1102 of Fig. 11. If the input speech is "Send mail to May", the type of speech is determined to be language information. Then, the unit of the language information

is further determined in step S1106.

[0072] If the speed of the user's speech is fast, the unit of the language information is determined to be a longer unit, for example, a word rather than a phoneme. If a history indicates that the recognition of words has not been very successful, a shorter unit, for example, a phoneme, may be used as long as it is determined that speech recognition can be performed with the highest accuracy by using units of phonemes. This determination may be automatically made by the apparatus so as to improve the recognition rate. Alternatively, if it is desired by the user that recognition be made in units of words, the user may set the unit for recognition.

[0073] If it is determined in step S1106 that the unit of the language information is, for example, a word, the word-recognition processing is performed in step S1103. The flow chart of this processing is shown in Fig. 17. In step S1701, speech recognition is performed in units of words by using the speech recognition unit. More specifically, in response to the input "Send mail to May", the word recognition is implemented by selecting the word "send" closest to the input sound information from the word dictionary provided for the speech recognition unit. In step S1702, the word "send" determined in the word recognition in step S1701 is obtained together with the similarity between "send" contained in the dictionary and the input speech (for example, 92% similarity). It is then determined in step S1703 whether the result of the word recognition is to be used.

[0074] Referring back to Fig. 11, if it is determined in step S1106 that the type of speech is a syllable, syllable-recognition processing is executed in step S1104. The flow chart of this processing is shown in Fig. 18. In step S1801, speech recognition is conducted in units of syllables by using the speech recognition unit. In this example, the syllable recognition in units of syllables is implemented by selecting the top N syllables similar to the input sound information from the syllable dictionary provided for the speech recognition unit. In step S1802, the result of syllable recognition determined in step S1801 is acquired together with the similarity between the top N syllables and the information contained in the dictionary. In step S1803, the syllable determined in step S1802 is then recognized as the word "send" which can be handled by the system, and the similarity between the whole word and the input speech is output. The details of this processing are indicated by the flow chart of Fig. 19. Upon performing recognition processing in units of syllables, "se" is acquired in step S1901. Then, in step S1902, a suitable word that matches the syllable "se" is determined by using the result obtained in step S1901. It is also determined in step S1902 whether the result of the word recognition in step S1901 is to be employed.

[0075] Referring back to Fig. 11, if it is found in step S1106 that the type of speech is neither a word nor a syllable, the speech recognition suitable for the corresponding type is performed in step S1105. The type of speech which is neither a word nor a syllable may be a phoneme, which is shorter than a syllable, or a stereotyped phrase, such as a sentence, which is longer than a syllable. A plurality of types of units may be used for recognizing input information until a series of operations has been completed by the user. Alternatively, only one type of unit may be used.

[0076] Referring back to Fig. 10, the result obtained upon recognizing the input information in step S1001 is processed in step S1002. More specifically, it is determined in step S1703 of Fig. 17 and in step S1902 of Fig. 19 whether the recognition result is to be finally accepted as the input by the user. The detailed process is shown in Fig. 37.

[0077] In step S3701, processing for determining the recognition result is executed. More specifically, it is determined in step S3701 whether the recognition result is to be accepted, for example, by providing a threshold for the similarity of the speech recognition. It is now assumed that the threshold similarity is set to be 80%. If the recognition result is "send: 85% similarity", it is determined in step S3702 that the recognition result is to be accepted, and it is notified in step S3703 that the recognition result "send" has been accepted. Conversely, if the recognition result is "send: 70% similarity", it is determined in step S3702 that the recognition result is to be rejected, and it is reported in step S3704 that the recognition result has been rejected, so that a subsequent user's input is ready to be processed.

[0078] Referring back to Fig. 8, after recognizing various types of input information in step S805, the recognition result "send" is analyzed in step S806. In step S806, the processing is executed by primarily analyzing the concept of the word "send". The flow of this processing is schematically shown in Fig. 9. The recognition result is used for predicting a subsequent object to be recognized in step S904. The prediction for the subsequent object may be made in step S904 by using not only the recognition result, but also an analysis of the recognition result (step S901), a determination of the correctness of the result (step S902), and a correction of the result (step S903). It is not essential that all the processing in steps S901 through S903 be performed. Alternatively, only an analysis of the recognition result or only a correction of the result may be performed. Alternatively, a combination of a determination of the correctness of the result and a correction of the result, or a combination of an analysis of the recognition result and a determination of the correctness of the result may be performed. In step S905, according to the prediction generated in step S904, the prediction is updated for subsequent information to be recognized.

[0079] An analysis of the recognition result in step S806 is more specifically shown by the flow chart of Fig. 38. In step S3801, a category search is conducted on the recognition result "send" so as to acquire information, such as the attribute of the word "send" and the context of the currently activated words. Subsequently, in step S3802, the language knowledge of the word

"send" is checked from the language dictionary, such as that shown in Fig. 29, and in step S3803, the concept knowledge is checked from the concept dictionary, such as that shown in Fig. 30, thereby obtaining the corresponding information. In step S3804, the operation and the object of the word "send" are then checked from the domain of the apparatus. Thereafter, in step S3805, the rules of the system concerning the input of the word "send" are checked from a rule table, such as that illustrated in Fig. 31. As a consequence, in step S3806, semantic analyses are comprehensively conducted by employing the above knowledge.

[0080] More specifically, the semantic analyses in step S3806 are implemented by conducting concept analyses in step S3901 of Fig. 39 by utilizing the knowledge concerning the word "send" obtained in steps S3802 through S3805. In step S3902, by referring to the language dictionary, such as that shown in Fig. 29, it is found that "send" is a verb, which is followed by an object or the name of an apparatus. In step S3803, by referring to the concept dictionary, such as that shown in Fig. 30, it is found that "send" represents a physical transfer. In step S3805, by checking the rule table, such as that shown in Fig. 31, it is found that the object of "send" is mail or a document. Upon conducting the concept analyses in step S3901, it is determined in step S3902 whether the recognition result "send" obtained through speech recognition is correct in terms of semantics and common sense. A determination is also made in step S3902 as to whether the recognition result "send" satisfies the prediction set in response to the current input.

[0081] If the domain is represented by a printer, the verb "send" may be included in the initial prediction. It is thus determined in step S3902 that the verb "send" satisfies the initial prediction. Then, in step S3903, a prediction is made on a subsequent input, which is preceded by "send", by utilizing various types of knowledge in the system. In this case, it is predicted from the word "send" that the user is likely to specify "sending what", and that a subsequent input is likely to be an object. The prediction set as described above is updated in step S3904.

[0082] More specifically, in step S3904, the word dictionary provided for the speech recognition unit may be updated. Then, upon the concept analyses, the processing to be executed by the system in accordance with the user's purpose is determined in step S3905. In this case, since the system cannot execute the processing merely by the word "send", it waits for a subsequent information to be input according to the determination of the analysis result made in step S807.

[0083] Upon receiving speech "mail" from the user, the process proceeds in a manner similar to the above process. It is now assumed that a recognition result "mail" instead of "mail" be returned after conducting the speech recognition processing based on the prediction that a subsequent input may be an object, i.e., a noun. In the processing for determining the recognition result in step S3701, the semantic analyses are conducted in step S3806 by using the knowledge checked in steps S3801 through S3805. Upon checking the domain knowledge in step S3804, it can be considered that "mail" is not suitable as the input word. Then, it is determined in step S3902 that the result "mail" should be corrected. The flow then proceeds to step S3906.

[0084] In order to re-determine the recognition result, a result re-determining processing is performed in step S4001 of Fig. 40. This processing is more specifically indicated by the flow chart of Fig. 41. In step S4101, the recognition result obtained so far is re-determined. If it is determined in step S4102 that the result should be corrected, the prediction for the previous input is re-generated in step S4103. In step S4104, the recognized word is re-determined. In this case, no corrections are made on the recognition result "send" since it is not necessary. Thus, a prediction for the current input is re-generated in step S4002 of Fig. 40 while the recognition result "send" remains the same. The re-generated prediction is then updated in step S4003, and the current input is again recognized by utilizing another type of knowledge in step S4004.

[0085] The updating of the prediction in step S4003 is to merge the newly generated prediction into the previous prediction. Accordingly, even after performing this updating operation, the number of predictions is not increased, which would otherwise generate more matches. On the contrary, the prediction becomes more precise to restrict the number of possible words. For example, the current prediction is updated in step S4003 by a prediction that "mail" is more likely to be input than "mail" by considering the domain of the system. Upon re-examining the current recognition result in step S4004, it is determined that "mail" is more suitable than "male".

[0086] As discussed above, upon completion of recognizing the words "Send mail to May", it is determined in step S807 of Fig. 8 that the system should execute the processing. Then, in step S808, a command indicating "Send mail to May" is executed. During execution, May's mail address is checked in an address book provided for the system to ensure that mail is appropriately sent. After performing step S808, if it is determined that a response should be returned to the user, output processing is executed in step S809. Fig. 28 is a flow chart illustrating the process of generating a response to the user. More specifically, in step S2801, the status of the execution result is acquired. Then, in step S2802, the response to be output to the user is analyzed. In step S2803, a response to the user is actually generated by utilizing the knowledge of the system, and in step S2804, an actual sentence is created. In step S2805, the optimal output method is selected, and in step S2806, the response is output to the user. For example, a confirmation message, such as "Mail has been sent to May", may be output to the user in sound by perform-

ing speech synthesizing.

Second Embodiment

[0087] A description is given below of the flow of the processing performed on a speech input, such as "Send mail to May" when a syllable dictionary, a word dictionary, a language dictionary, and a concept dictionary are provided. as illustrated in Figs. 25, 26, 29, and 30, respectively. A parameter setting procedure in the recognition processing is also discussed. It is now assumed that M number of words are registered in the word dictionary, and each word is divided into R[I] number of syllables, as shown in Fig. 32.

[0088] Upon the start of the system, the screen, such as that shown in Fig. 24, appears. The parameter setting procedure is indicated by the flow chart of Fig. 23. In step S2301, parameters for setting the speech recognition engine, the Garbage level, the Rejection Time, and the Minimum Speech Duration, etc. are indicated. In step S2302, a reference of the similarity is indicated as a condition for determining the recognition result. Thereafter, in step S2303, the content of the currently activated context is indicated.

[0089] When the speech recognition method is employed, in which words are inferred from the result of syllable recognition in response to a speech input "Send mail to May", the following processing is performed. Upon receiving sound information "send", the sound information is first processed in step S1101 of Fig. 11. If it is determined in step S1102 that the type of processed information is a syllable, syllable-recognition processing is performed in step S1104. Then, speech recognition is conducted in units of syllables by using the speech recognition engine in step S1801 of Fig. 18, thereby obtaining the top N syllables as the recognition results. Referring back to Fig. 23, in step S2304, the recognition results obtained by the speech recognition engine and the similarity are acquired, and in step S2305, the obtained information, i.e., the recognition results in units of syllables and the similarity, is indicated according to the highest degree. In step S2306, the recognition results output from the speech recognition engine are determined, and the results are output after a result of the determination processing.

[0090] In the determination processing in step S2306, the results of the speech recognition engine are obtained in step S1802 of Fig. 18. In response to a speech input, i.e., "send", the result, such as that shown in Fig. 33, is obtained. By utilizing this result, syllable-to-word matching processing is performed in step S1803. More specifically, in step S1901, matching processing is performed to determine a suitable word from the syllables. In this embodiment, a dynamic programming (DP) matching method is used as the syllable-to-word matching method. In this DP matching method, matching is made between the word dictionary and the input speech according to the algorithm illustrated in Fig. 34.

[0091] The flow of the DP matching processing is shown in Fig. 20. In step S2001, the number of top syllables N and the number of syllables T recognized so far (recognition time) are acquired. Fig. 35 reveals that N is 5 and T is 0. Then, in step S2002, the number of words M registered in the system and the number of syllables R[i] forming each word are acquired. Fig. 36 reveals that M is 3, and R[1], R[2], and R[3] are 2, 2, and 4, respectively. In step S2003, i is set to be 1. Then, while the condition, i.e., i ≤ M, set in step S2004 is satisfied, the following processing is repeated. When it is determined in step S2004 that I is 1, the process proceeds to step S2005 in which the dictionary word W[i], i.e., "send", is obtained. In step S2006, j is set to be 1. Then, while the condition, i.e., j < R[1](=2), is satisfied, the processing for obtaining syllables forming "send" is repeated. Then, in step S2008, S[1][1] = "se" is obtained.

[0092] Subsequently, in step S2101 of Fig. 21, k is set to be 1. While the condition, i.e., k ≤ N(=5), is met in step S2102, it is determined whether S[1][1] is returned as a syllable that may match the input information. When it is determined in step S2102 that k is 1, the recognition result C[k] and the similarity A[k] are acquired in step S2103, resulting in C[1] = "nd" and A[1] = 60.4. When it is determined in step S2105 that S[1][1] ≠ C[1], k is incremented by one in step S2104, and a subsequent syllable that may match the input information is obtained and determined. When C[3] = "se" and A[3] = 38.9, the outcome of step S2105 is yes. Thus, the process proceeds to step S2106 in which the similarity D[1][1] of the syllable S[1][1] = A[3] is set to be 38.9. Then, in step S2108, the similarity CD[1][1] of the word W[1] = "send" is calculated.

[0093] Referring to Fig. 22, it is then determined in step S2201 whether T is 0. In this case, since T is 0, the optimal path is calculated in step S2211 according to the equations illustrated in Fig. 34. That is, it is set in step S2211 that P1 = 1, P2 = 2 * 60.4 = 120.8, and P3 = 0. It is thus determined in step S2212 that P2 = 120.8 is the optimal path. In step S2213, the cumulative similarity and the cumulative path are calculated.

[0094] In this case, since the optimal path is 2, the cumulative similarity CD[1][1] is 120.8 and the cumulative path is 2. Referring back to Fig. 21, j is incremented by one, i.e., j = 2, in step S2109. Then, the recognition result CD[1] = "nd" and the similarity A[1] = 61.0 are acquired in step S2103. Accordingly, the determination of step S2105 becomes true. The above-described calculations are then made in step S2108, and the result CD [1][2] = 122.0 is obtained. Thereafter, j is incremented by one, i.e., j = 3 in step S2109, and the determination of step S2007 becomes false. Accordingly, i is incremented by one, i = 2, in step S2009. In this manner, the processing is repeated until i becomes 3. As a result, the similarity of words CD[1][1] = 120.8, CD[1][2] = 122.0, CD[2][1] = 107.4, CD[2][2] = 41.2, CD[3][1] = 58.2, CD[3][2] = 0, CD[3][3] = 0, and CD[3][4] = 0 are obtained. When i becomes 4 in step S2009, the deter-

mination of step S2004 becomes false. Thus, the processing is completed, and the word-recognition result is subsequently determined in step S1902 of Fig. 19.

[0095] In step S1902, the result of syllable-to-word matching obtained when T is 0 is determined. According to the calculations conducted as described above, the exact word that may match the input sound has not yet been determined, and the system waits for a subsequent input. Upon receiving the subsequent input, the type of sound is determined in step S1102 of Fig. 11. In this case, since the previous input is a syllable and has not yet been recognized as a word, it is determined that the type of subsequent input may be again a syllable. The syllable-recognition processing is then performed in step S1104, and the syllable recognition results, such as those illustrated in Fig. 35, are returned in step S1802.

[0096] The recognition results are thus obtained in step S1802, and syllable-to-word matching is conducted in step S1803. In step S2001 of Fig. 20, the number of top syllables N obtained as the recognition results and the recognition time T are acquired. That is, N=1 and T=1 are obtained. As in the case of the processing executed when T=0, steps S2002 through S2109 are executed. As a result, the calculation results CD[1][1] = 120.8, CD[1][2] = 322.0, CD[2][1] = 107.4, CD[2][2] = 41.2, CD[3][1] = 58.2, CD[3][2] = 0, CD[3][3] = 0, and CD[3][4] = 0 are obtained. When i becomes 4 in step S2009, the determination of step S2004 becomes false, and the process proceeds to step S1902 in which the word-recognition result is determined.

[0097] According to the determination in step S1902, the word "send" is set to be the recognition result obtained by performing the recognition processing in step S805 of Fig. 8. It is determined in step S807 that the word "send" is to be accepted after analyzing the processed result in step S806.

[0098] According to the foregoing description, the speech input "Send mail to May" is first processed, and mail is transmitted to May, and finally, a response is output to the user. The whole processing is then completed.

[0099] Predictions conducted in units of a short speech unit, for example, a syllable, are effective particularly when only part of the whole speech was recognized. For example, even if input sound forming a word was not completely recognized, highly precise recognition can be expected if predictions made by utilizing the knowledge base are used in combination with occurrence probability of each unit and the inter-state transition probability.

[0100] In outputting the recognized result, the determination of step S2307 is made based on the analysis of the processed result in step S806. If it is determined in step S2307 that the recognition result "send" is accepted, it is output as the final recognition result in step S2308. Similarly, the aforementioned processing is performed on a subsequent input. If there is a change of the parameters on the screen shown in Fig. 24, the newly set parameters are made valid, and determinations are made by using the new parameters.

Third Embodiment

[0101] According to the first and second embodiments, English-language speech recognition is performed. However, the Japanese language may be recognized based on predictions, and the corresponding processing may be appropriately executed by conducting concept analysis. In this case, language information is provided as a dictionary, and concept analyses are not dependent upon the type of language. Thus, differences in languages, such as English and Japanese, do not influence speech recognition by utilizing concept analyses.

Fourth Embodiment

[0102] According to the prediction technique shown in Fig. 13, not only speech information, which is likely to be input, but also information which is unlikely to be input, is predicted. Thus, the fact that exactly the same information is not input consecutively may be used for predictions so as to eliminate information which is unlikely to be input, thereby enhancing the recognition efficiency.

[0103] As is seen from the foregoing description, the present invention offers the advantage of improving the recognition accuracy by performing speech recognition based on predictions.

[0104] The present invention may be applied to a single apparatus or to a system formed of a plurality of apparatuses.

[0105] In another embodiment of the present invention, software program code for implementing the above-described functions may be supplied to an apparatus or a system, and a computer within the apparatus or the system may read the program code stored in a storage medium and execute it, so that the above-described functions can be implemented.

[0106] The function of the foregoing embodiment can be implemented not only by running the program code read by the computer, but also by executing the processing by, for example, an operating system (OS), running in the computer according to the instructions of the program code.

[0107] According to the above-described modifications, a storage medium storing the program code constitutes the present invention.

[0108] Although the present invention has been described in its preferred form with a certain degree of particularity, many apparently widely different embodiments of the invention can be made without departing from the scope of the claims. It is to be understood that the invention is not limited to the specific embodiments thereof except as defined in the appended claims.

Claims

1. A speech information processing apparatus comprising:

   a context base (207) for storing prediction information representing a category of words to be subsequently recognised;
   recognition means (601) for recognising a word based on the category of words represented by the prediction information stored in said context base (207);
   a knowledge base (208) for storing knowledge concerning a category of speech information;
   prediction means for predicting the category of words to be subsequently recognised based on at least one previously recognised word by referring to the knowledge stored in said knowledge base (208); and
   updating means for updating the prediction information stored in said context base (207) based on the category of words obtained by said prediction means;

   characterised in that:

   the apparatus further comprises determination means for determining whether or not input sound information is language information by referring to the knowledge stored in said knowledge base (208); and **in that**
   said recognition means (601) is arranged to perform the recognition on the input sound information when said determination means determines that the input sound information is language information.

2. A speech information processing apparatus according to claim 1, wherein said prediction means is operable to further predict a category of words which is less likely to be subsequently recognised and wherein said recognition means (601) is arranged to reject the category of words which is less likely to be subsequently recognised.

3. A speech information processing apparatus according to claim 1 or 2, further comprising correctness determination means for determining the correctness of a recognition result obtained by said recognition means (601), and wherein said prediction means is arranged to perform a prediction based on a determination result obtained by said correctness determination means.

4. A speech information processing apparatus according to claim 3, further comprising correction means for correcting the recognition result based on a determination result obtained by said correctness determination means, and wherein said prediction means is arranged to perform a re-prediction based on the recognition result corrected by said correction means.

5. A speech information processing apparatus according to claim 4, wherein said correction means is arranged to correct a new recognition result by referring to the knowledge stored in said knowledge base (208) based on a previous recognition result.

6. A speech information processing apparatus according to any preceding claim, further comprising sound determining means for determining whether or not sound information is speech information produced by a human, and wherein said determination means is arranged to perform a determination on the input sound information determined by said sound determining means.

7. A speech information processing apparatus according to claim 6, wherein said sound determining means is arranged to distinguish human speech from a mechanical sound based on a frequency difference.

8. A speech information processing apparatus according to any preceding claim, wherein said recognition means (601) is arranged to recognise a word in units of one of: words, syllables and phonemes.

9. A speech information processing apparatus according to claim 8, wherein said prediction means is arranged to predict a word including one of a syllable and a phoneme to be subsequently recognised when said recognition means (601) is arranged to recognise words in units of one of: syllables and phonemes, respectively, and wherein said prediction means is arranged to predict one of: the syllable or the phoneme to be subsequently recognised based on the category of words represented by the prediction information.

10. A speech information processing apparatus according to claim 8 or 9, further comprising selection means for selecting the unit for recognition according to whether or not a previous recognition result was successfully obtained.

11. A speech information processing apparatus according to any preceding claim, wherein said prediction means is arranged to predict the category of words to be subsequently recognised based on a previous operation.

12. A speech information processing apparatus according to claim 11, wherein said prediction means

is arranged to predict a subsequent operation based on the previous operation and to predict the category of words to be subsequently recognised based on the predicted subsequent operation.

13. A speech information processing apparatus according to claim 11 or 12, wherein said prediction means is arranged to predict the category of words to be initially recognised based on the previous operation.

14. A speech information processing method comprising:

a recognition step (S805) of recognising a word based on a category of words to be subsequently recognised which is represented by prediction information stored in a context base (207);
a prediction step (S904) of predicting the category of words to be subsequently recognised based on at least one previously recognised word by referring to knowledge concerning a category of speech information stored in a knowledge base (208); and
an updating step (S905) of updating the prediction information stored in said context base (207) based on the category of words obtained in said prediction step;

characterised by a determination step (S1108) of determining whether or not input sound information is language information by referring to the knowledge stored in said knowledge base (208); and in that said recognition step (S805) performs the recognition on the input sound information when the input sound information is determined to be language information in said determination step (S1108).

15. A speech information processing method according to claim 14, wherein said prediction step (S904) further predicts a category of words which is less likely to be subsequently recognised and wherein said recognition step (S805) rejects the category of words which is less likely to be subsequently recognised.

16. A speech information processing method according to claim 14 or 15, further comprising a correctness determination step (S902) of determining the correctness of a recognition result obtained in said recognition step (S805), and wherein said prediction step (S904) performs a prediction based on a determination result obtained in said correctness determination step.

17. A speech information processing method according to claim 16, further comprising a correction step (S903) of correcting the recognition result based on a determination result obtained in said correctness determination step (S902), and wherein said prediction step (S904) performs a re-prediction based on the recognition result corrected in said correction step (S903).

18. A speech information processing method according to claim 17, wherein said correction step (S903) corrects a new recognition result by referring to the knowledge stored in said knowledge base (208) based on a previous recognition result.

19. A speech information processing method according to any of claims 14 to 18, further comprising a sound determining step (S1202) of determining whether or not sound information is speech information produced by a human, and wherein said determination step performs a determination on the input sound information determined in said sound determining step.

20. A speech information processing method according to claim 19, wherein said sound determining step (S1108) distinguishes human speech from a mechanical sound based on a frequency difference.

21. A speech information processing method according to any of claims 14 to 20, wherein said recognition step (S805) recognises a word in units of one of: words, syllables and phonemes.

22. A speech information processing method according to claim 21, wherein said prediction step (S904) predicts a word including one of: a syllable and a phoneme to be subsequently recognised when said recognition step (S805) uses one of: syllables and phonemes, respectively, as a unit of recognition, and predicts one of the syllable or the phoneme to be subsequently recognised based on the category of words represented by the prediction information.

23. A speech information processing method according to claim 21 or 22, further comprising a selection step of selecting the unit of recognition according to whether or not a previous recognition result was successfully obtained.

24. A speech information processing method according to any of claims 14 to 23, wherein said prediction step (S904) predicts the category of words to be subsequently recognised based on a previous operation.

25. A speech information processing method according to claim 24, wherein said prediction step (S904) predicts a subsequent operation based on the previous

operation and predicts the category of words to be subsequently recognised based on the predicted subsequent operation.

26. A speech information processing method according to claim 24 or 25, wherein said prediction step predicts the category of words to be initially recognised based on the previous operation.

27. A storage medium storing processor implementable instructions for controlling a processor to implement the method of any of claims 14 to 26.

28. Processor implementable instructions for controlling a processor to implement the method of any of claims 14 to 26.

Patentansprüche

1. Sprachinformationsverarbeitungsvorrichtung mit:

einer Kontextbasis (207) zum Speichern von eine Kategorie von anschließend zu erkennenden Wörtern darstellenden Vorhersageinformationen;
einer Erkennungseinrichtung (601) zum Erkennen eines Worts basierend auf der durch die in der Kontextbasis (207) gespeicherten Vorhersageinformationen dargestellten Kategorie von Wörtern;
einer Wissensbasis (208) zum Speichern von eine Kategorie von Sprachinformationen betreffendem Wissen;
einer Vorhersageeinrichtung zum Vorhersagen der Kategorie von anschließend zu erkennenden Wörtern basierend auf zumindest einem vorher erkannten Wort mit Bezug auf das in der Wissensbasis (208) gespeicherte Wissen; und
einer Aktualisierungseinrichtung zum Aktualisieren der in der Kontextbasis (207) gespeicherten Vorhersageinformationen basierend auf der durch die Vorhersageeinrichtung erhaltenen Kategorie von Wörtern;

dadurch gekennzeichnet, dass:

die Vorrichtung ferner eine Bestiminungseinrichtung zum Bestimmen, ob eingegebene Toninformationen Sprachinformationen sind oder nicht, mit Bezug auf das in der Wissensbasis (208) gespeicherte Wissen umfasst; und dass

die Erkennungseinrichtung (601) zum Durchführen der Erkennung bei den eingegebenen Toninformationen, wenn die Bestimmungseinrichtung bestimmt, dass die eingegebenen Toninformationen Sprachinformationen sind, ausgelegt ist.

2. Sprachinformationsverarbeitungsvorrichtung nach Anspruch 1, wobei die Vorhersageeinrichtung dazu betreibbar ist, ferner eine Kategorie von Wörtern, deren anschließende Erkennung weniger wahrscheinlich ist, vorherzusagen, und wobei die Erkennungseinrichtung (601) zum Ablehnen der Kategorie von Wörtern, deren anschließende Erkennung weniger wahrscheinlich ist, ausgelegt ist.

3. Sprachinformationsverarbeitungsvorrichtung nach Anspruch 1 oder 2, ferner mit einer Richtigkeitsbestimmungseinrichtung zum Bestimmen der Richtigkeit eines durch die Erkennungseinrichtung (601) erhaltenen Erkennungsergebnisses, und wobei die Vorhersageeinrichtung zum Durchführen einer Vorhersage basierend auf einem durch die Richtigkeitsbestimmungseinrichtung erhaltenen Bestimmungsergebnis ausgelegt ist.

4. Sprachinformationsverarbeitungsvorrichtung nach Anspruch 3, ferner mit einer Berichtigungseinrichtung zum Berichtigen des Erkennungsergebnisses basierend auf einem durch die Richtigkeitsbestimmungseinrichtung erhaltenen Bestimmungsergebnis, und wobei die Vorhersageeinrichtung zum Durchführen einer Neuvorhersage basierend auf dem durch die Berichtigungseinrichtüng berichtigten Erkennungsergebnis ausgelegt ist.

5. Sprachinformationsverarbeitungsvorrichtung nach Anspruch 4, wobei die Berichtigungseinrichtung zum Berichtigen eines neuen Erkennungsergebnisses mit Bezug auf das in der Wissensbasis (208) gespeicherte Wissen basierend auf einem vorhergehenden Erkennungsergebnis ausgelegt ist.

6. Sprachinformationsverarbeitungsvorrichtung nach einem der vorstehenden Ansprüche, ferner mit einer Tonbestimmungseinrichtung zum Bestimmen, ob Toninformationen durch einen Menschen erzeugte Sprachinformationen sind oder nicht, und wobei die Bestimmungseinrichtung zum Durchführen einer Bestimmung bei den durch die Tonbestimmungseinrichtung bestimmten eingegebenen Toninformationen ausgelegt ist.

7. Sprachinformationsverarbeitungsvorrichtung nach Anspruch 6, wobei die Tonbestimmungseinrichtung zum Unterscheiden von menschlicher Sprache von einem mechanischen Ton basierend auf einem Frequenzunterschied ausgelegt ist.

8. Sprachinformationsverarbeitungsvorrichtung nach einem der vorstehenden Ansprüche, wobei die Erkennungseinrichtung (601) zum Erkennen eines Worts in Einheiten von einem der Nachstehenden

ausgelegt ist: Wörter, Silben und Phoneme.

9. Sprachinformationsverarbeitungsvorrichtung nach Anspruch 8, wobei die Vorhersageeinrichtung zum Vorhersagen eines Worts einschließlich einer Silbe oder eines Phonems, die oder das anschließend zu erkennen ist, ausgelegt ist, wenn die Erkennungseinrichtung (601) zum Erkennen von Wörtern in Einheiten eines der Nachstehenden ausgelegt ist: Silben bzw. Phoneme, und wobei die Vorhersageeinrichtung zum Vorhersagen eines der Nachstehenden ausgelegt ist: der Silbe oder des Phonems, die oder das anschließend basierend auf der durch die Vorhersageinformationen dargestellten Kategorie von Wörtern zu erkennen ist.

10. Sprachinformationsverarbeitungsvorrichtung nach Anspruch 8 oder 9, ferner mit einer Auswahleinrichtung zum Auswählen der Einheit zur Erkennung gemäß dem, ob ein vorhergehendes Erkennungsergebnis erfolgreich erhalten wurde oder nicht.

11. Sprachinformationsverarbeitungsvorrichtung nach einem der vorstehenden Ansprüche, wobei die Vorhersageeinrichtung zum Vorhersagen der Kategorie von anschließend zu erkennenden Wörtern basierend auf einer vorhergehenden Operation ausgelegt ist.

12. Sprachinformationsverarbeitungsvorrichtung nach Anspruch 11, wobei die Vorhersageeinrichtung zum Vorhersagen einer nachfolgenden Operation basierend auf der vorhergehenden Operation und zum Vorhersagen der Kategorie von anschließend zu erkennenden Wörtern basierend auf der vorhergesagten nachfolgenden Operation ausgelegt ist.

13. Sprachinformationsverarbeitungsvorrichtung nach Anspruch 11 oder 12, wobei die Vorhersageeinrichtung zum Vorhersagen der Kategorie von anfänglich zu erkennenden Wörtern basierend auf der vorhergehenden Operation ausgelegt ist.

14. Sprachinformationsverarbeitungsverfahren mit:

einem Erkennungsschritt (S805) des Erkennens eines Worts basierend auf einer Kategorie von anschließend zu erkennenden Wörtern, die durch in einer Kontextbasis (207) gespeicherte Vorhersageinformationen dargestellt wird;
einem Vorhersageschritt (S904) des Vorhersagens der Kategorie von anschließend zu erkennenden Wörtern basierend auf zumindest einem vorher erkannten Wort mit Bezug auf eine Kategorie von Sprachinformationen betreffendes Wissen, das in einer Wissensbasis (208) gespeichert ist; und
einem Aktualisierungsschritt (S905) des Aktualisierens der in der Kontextbasis (207) gespeicherten Vorhersageinformationen basierend auf der in dem Vorhersageschritt erhaltenen Kategorie von Wörtern;

gekennzeichnet durch einen Bestimmungsschritt (S1108) des Bestimmens, ob eingegebene Toninformationen Sprachinformationen sind oder nicht, mit Bezug auf das in der Wissensbasis (208) gespeicherte Wissen; und **dadurch**, dass der Erkennungsschritt (S805) die Erkennung bei den eingegebenen Toninformationen durchführt, wenn in dem Bestimmungsschritt (S1108) bestimmt wird, dass die eingegebenen Toninformationen Sprachinformationen sind.

15. Sprachinformationsverarbeitungsverfahren nach Anspruch 14, wobei der Vorhersageschritt (S904) ferner eine Kategorie von Wörtern, deren anschließende Erkennung weniger wahrscheinlich ist, vorhersagt, und wobei der Erkennungsschritt (S805) die Kategorie von Wörtern, deren anschließende Erkennung weniger wahrscheinlich ist, ablehnt.

16. Sprachinformationsverarbeitungsverfahren nach Anspruch 14 oder 15, ferner mit einem Richtigkeitsbestimmungsschritt (S902) des Bestimmens der Richtigkeit eines in dem Erkennungsschritt (S805) erhaltenen Erkennungsergebnisses, und wobei der Vorhersageschritt (S904) eine Vorhersage basierend auf einem in dem Richtigkeitsbestimmungsschritt erhaltenen Bestimmungsergebnis durchführt.

17. Sprachinformationsverarbeitungsverfahren nach Anspruch 16, ferner mit einem Berichtigungsschritt (S903) des Berichtigens des Erkennungsergebnisses basierend auf einem in dem Richtigkeitsbestimmungsschritt (S902) erhaltenen Bestimmungsergebnis, und wobei der Vorhersageschritt (S904) eine Neuvorhersage basierend auf dem in dem Berichtigungsschritt (S903) berichtigten Erkennungsergebnis durchführt.

18. Sprachinformationsverarbeitungsverfahren nach Anspruch 17, wobei der Berichtigungsschritt (S903) ein neues Erkennungsergebnis mit Bezug auf das in der Wissensbasis (208) gespeicherte Wissen basierend auf einem vorhergehenden Erkennungsergebnis berichtigt.

19. Sprachinformationsverarbeitungsverfahren nach einem der Ansprüche 14 bis 18, ferner mit einem Tonbestimmungsschritt (S1202) des Bestimmens, ob Toninformationen durch einen Menschen erzeugte Sprachinformationen sind oder nicht, und wobei der Bestimmungsschritt eine Bestimmung

bei den in dem Tonbestimmungsschritt bestimmten eingegebenen Tonlnformationen durchführt.

20. Sprachinformationsverarbeituhgsverfahren nach Anspruch 19, wobei der Tonbestimmungsschritt (S1108) basierend auf einem Frequenzunterschied menschliche Sprache von einem mechanischen Ton unterscheidet.

21. Sprachinformationsverarbeitungsverfahren nach einem der Ansprüche 14 bis 20, wobei der Erkennungsschritt (S805) ein Wort in Einheiten eines der Nachstehenden erkennt: Wörter, Silben und Phoneme.

22. Sprachinformationsverarbeitungsverfahren nach Anspruch 21, wobei der Vorhersageschritt (S904) ein Wort einschließlich eines der Nachstehenden vorhersagt: einer Silbe und eines Phonems, die oder das anschließend zu erkennen ist, wenn der Erkennungsschritt (S805) eines der Nachstehenden verwendet: Silben bzw. Phoneme als eine Einheit der Erkennung, und die Silbe oder das Phonem, die oder das anschließend zu erkennen ist, basierend auf der durch die Vorhersageinformationen dargestellten Kategorie von Wörtern vorhersagt.

23. Sprachinformationsverarbeitungsverfahren nach Anspruch 21 oder 22, ferner mit einem Auswahlschritt des Auswählens der Einheit der Erkennung gemäß dem, ob ein vorhergehendes Erkennungsergebnis erfolgreich erhalten wurde oder nicht.

24. Sprachinformationsverarbeitungsverfahren nach einem der Ansprüche 14 bis 23, wobei der Vorhersageschritt (S904) die Kategorie von anschließend zu erkennenden Wörtern basierend auf einer vorhergehenden Operation vorhersagt.

25. Sprachinformationsverarbeitungsverfahren nach Anspruch 24, wobei der Vorhersageschritt (S904) eine nachfolgende Operation basierend auf der vorhergehenden Operation vorhersagt und die Kategorie von anschließend zu erkennenden Wörtern basierend auf der vorhergesagten nachfolgenden Operation vorhersagt.

26. Sprachinformationsverarbeitungsverfahren nach Anspruch 24 oder 25, wobei der Vorhersageschritt die Kategorie von anfänglich zu erkennenden Wörtern basierend auf der vorhergehenden Operation vorhersagt.

27. Speichermedium, das durch eine Verarbeitungseinrichtung ausführbare Anweisungen zum Steuern einer Verarbeitungseinrichtung zum Ausführen des Verfahrens nach einem der Ansprüche 14 bis 26 speichert.

28. Durch eine Verarbeitungseinrichtung ausführbare Anweisungen zum Steuern einer Verarbeitungseinrichtung zum Ausführen des Verfahrens nach einem der Ansprüche 14 bis 26.

Revendications

1. Appareil de traitement d'informations vocales, comportant :

une base (207) de contexte destinée à stocker une information de prédiction représentant une catégorie de mots devant être ensuite reconnue ;
un moyen (601) de reconnaissance destiné à reconnaître un mot basé sur la catégorie de mots représentée par l'information de prédiction stockée dans ladite base (207) de contexte ;
une base (208) de connaissances destinée à stocker des connaissances concernant une catégorie d'informations vocales ;
un moyen de prédiction destiné à prédire la catégorie de mots devant être ensuite reconnue sur la base d'au moins un mot précédemment reconnu en se référant aux connaissances stockées dans ladite base (208) de connaissances ; et
un moyen de mise à jour destiné à mettre à jour l'information de prédiction stockée dans ladite base (207) de contexte sur la base de la catégorie de mots obtenue par ledit moyen de prédiction ;

caractérisé en ce que :

l'appareil comporte en outre un moyen de détermination destiné à déterminer si une information sonore d'entrée est ou non une information de langage en se référant aux connaissances stockées dans ladite base (208) de reconnaissances ; et

en ce que

ledit moyen (601) de reconnaissance est agencé pour effectuer la reconnaissance de l'information sonore d'entrée lorsque ledit moyen de détermination détermine que l'information sonore d'entrée est une information de langage.

2. Appareil de traitement d'informations vocales selon la revendication 1, dans lequel ledit moyen de prédiction peut être mis en oeuvre pour prédire en outre une catégorie de mots qui est moins susceptible d'être ensuite reconnue et dans lequel ledit moyen de reconnaissance (601) est agencé de fa-

çon à rejeter la catégorie de mots qui est moins susceptible d'être ensuite reconnue.

3. Appareil de traitement d'informations vocales selon la revendication 1 ou 2, comportant en outre un moyen de détermination d'exactitude destiné à déterminer l'exactitude d'un résultat de reconnaissance obtenu par ledit moyen de reconnaissance (601), et dans lequel ledit moyen de prédiction est agencé pour effectuer une prédiction sur la base d'un résultat de détermination obtenu par ledit moyen de détermination d'exactitude.

4. Appareil de traitement d'informations vocales selon la revendication 3, comportant en outre un moyen de correction destiné à corriger le résultat de la reconnaissance sur la base d'un résultat de détermination obtenu par ledit moyen de détermination d'exactitude, et dans lequel ledit moyen de prédiction est agencé pour effectuer une nouvelle prédiction sur la base du résultat de la reconnaissance corrigé par ledit moyen de correction.

5. Appareil de traitement d'informations vocales selon la revendication 4, dans lequel ledit moyen de correction est agencé pour corriger un nouveau résultat de reconnaissance en se référant aux connaissances stockées dans ladite base (208) de connaissances sur la base d'un résultat de reconnaissance précédent.

6. Appareil de traitement d'informations vocales selon l'une quelconque des revendications précédentes, comportant en outre un moyen de détermination de sons destiné à déterminer si une information sonore est ou non une information vocale produite par un être humain, et dans lequel ledit moyen de détermination est agencé pour effectuer une détermination sur l'information de sons d'entrée déterminée par ledit moyen de détermination de sons.

7. Appareil de traitement d'informations vocales selon la revendication 6, dans lequel ledit moyen de détermination de sons est agencé pour distinguer une parole humaine d'un son mécanique sur la base d'une différence de fréquence.

8. Appareil de traitement d'informations vocales selon l'une quelconque des revendications précédentes, dans lequel ledit moyen de reconnaissance (601) est agencé pour reconnaître un mot en unités de l'un de : mots, syllabes et phonèmes.

9. Appareil de traitement d'informations vocales selon la revendication 8, dans lequel ledit moyen de prédiction est agencé pour prédire un mot comprenant l'un d'une syllabe et d'un phonème devant être ensuite reconnu lorsque ledit moyen de reconnaissance (601) est agencé pour reconnaître des mots en unités de l'un de : syllabes et phonèmes, respectivement, et dans lequel ledit moyen de prédiction est agencé pour prédire l'un de : la syllabe ou le phonème devant être ensuite reconnu sur la base de la catégorie de mots représentée par l'information de prédiction.

10. Appareil de traitement d'informations vocales selon la revendication 8 ou 9, comportant en outre un moyen de sélection destiné à sélectionner l'unité pour la reconnaissance selon qu'un résultat de reconnaissance précédent a été obtenu avec succès ou non.

11. Appareil de traitement d'informations vocales selon l'une quelconque des revendications précédentes, dans lequel ledit moyen de prédiction est agencé pour prédire la catégorie de mots devant être ensuite reconnue sur la base d'une opération précédente.

12. Appareil de traitement d'informations vocales selon la revendication 11, dans lequel ledit moyen de prédiction est agencé pour prédire une opération suivante sur la base de l'opération précédente et pour prédire la catégorie de mots devant être ensuite reconnue sur la base de l'opération suivante prédite.

13. Appareil de traitement d'informations vocales selon la revendication 11 ou 12, dans lequel ledit moyen de prédiction est agencé pour prédire la catégorie de mots devant être reconnue initialement sur la base de l'opération précédente.

14. Procédé de traitement d'informations vocales comprenant :

une étape de reconnaissance (S805) consistant à reconnaître un mot sur la base d'une catégorie de mots devant être ensuite reconnue, qui est représentée par une information de prédiction stockée dans une basa (207) de contexte ;
une étape de prédiction (S904) consistant à prédire la catégorie de mots devant être ensuite reconnue sur la base d'au moins un mot précédemment reconnu en se référant à des connaissances concernant une catégorie d'informations vocales stockées dans une base (208) de connaissances ; et
une étape (S905) de mise à jour consistant à mettre à jour l'information de prédiction stockée dans ladite base (207) de contexte sur la base de la catégorie de mots obtenue dans ladite étape de prédiction ;

caractérisé par une étape de détermination

(S1108) consistant à déterminer si une information sonore d'entrée est ou non une information de langue en se référant aux connaissances stockées dans ladite base (208) de connaissances ; et en ce que ladite étape de reconnaissance (S805) effectue la reconnaissance sur l'information sonore d'entrée lorsque l'information sonore d'entrée est déterminée comme étant une information de langue dans ladite étape de détermination (S1108).

15. Procédé de traitement d'informations vocales selon la revendication 14, dans lequel ladite étape de prédiction (S904) prédit en outre une catégorie de mots moins susceptible d'être ensuite reconnue et dans lequel ladite étape de reconnaissance (S805) rejette la catégorie de mots moins susceptible d'être ensuite reconnue.

16. Procédé de traitement d'informations vocales selon la revendication 14 ou 15, comprenant en outre une étape (S902) de détermination d'exactitude consistant à déterminer l'exactitude d'un résultat de reconnaissance obtenu dans ladite étape de reconnaissance (S805), et dans lequel ladite étape de prédiction (S904) effectue une prédiction sur la base d'un résultat de détermination obtenu dans ladite étape de détermination d'exactitude.

17. Procédé de traitement d'informations vocales selon la revendication 16, comprenant en outre une étape de correction (S903) consistant à corriger le résultat de reconnaissance basé sur un résultat de détermination obtenu dans ladite étape (S902) de détermination d'exactitude, et dans lequel ladite étape de prédiction (S904) effectue une nouvelle prédiction basée sur le résultat de reconnaissance corrigé dans ladite étape de correction (S903).

18. Procédé de traitement d'informations vocales selon la revendication 17, dans lequel ladite étape de correction (S903) corrige un nouveau résultat de reconnaissance en se référant aux connaissances stockées dans ladite base (208) de connaissances sur la base d'un résultat de reconnaissance précédent.

19. Procédé de traitement d'informations vocales selon l'une quelconque des revendications 14 à 18, comprenant en outre une étape (S1202) de détermination de sons consistant à déterminer si une information sonore est ou non une information vocale produite par un être humain, et dans lequel ladite étape de détermination effectue une détermination sur l'information sonore d'entrée déterminée dans ladite étape de détermination de sons.

20. Procédé de traitement d'informations vocales selon la revendication 19, dans lequel ladite étape (S1108) de détermination de sons distingue une parole humaine d'un son mécanique sur la base d'une différence de fréquence.

21. Procédé de traitement d'informations vocales selon l'une quelconque des revendications 14 à 20, dans lequel ladite étape (S805) de reconnaissance reconnaît un mot en unités de l'un de : mots, syllabes et phonèmes.

22. Procédé de traitement d'informations vocales selon la revendication 21, dans lequel ladite étape de prédiction (S904) prédit un mot comprenant l'un de : une syllabe et un phonème devant être ensuite reconnus lorsque ladite étape de reconnaissance (S805) utilise l'un de : syllabes et phonèmes, respectivement, en tant qu'unité de reconnaissance, et prédit l'un de la syllabe ou du phonème devant être ensuite reconnu sur la base de la catégorie de mots représentée par l'information de prédiction.

23. Procédé de traitement d'informations vocales selon la revendication 21 ou 22, comprenant en outre une étape de sélection consistant à sélectionner l'unité de reconnaissance selon qu'un résultat de reconnaissance précédent a été obtenu avec succès ou non.

24. Procédé de traitement d'informations vocales selon l'une quelconque des revendications 14 à 23, dans lequel ladite étape de prédiction (S904) prédit la catégorie de mots devant être ensuite reconnue sur la base d'une opération précédente.

25. Procédé de traitement d'informations vocales selon la revendication 24, dans lequel ladite étape de prédiction (S904) prédit une opération suivante sur la base de l'opération précédente et prédit la catégorie de mots devant être ensuite reconnue sur la base de l'opération suivante prédite.

26. Procédé de traitement d'informations vocales selon la revendication 24 ou 25, dans lequel ladite étape de prédiction prédit la catégorie de mots devant être reconnue initialement sur la base de l'opération précédente.

27. Support de stockage stockant des instructions exécutables par un processeur pour commander un processeur afin qu'il exécute le procédé selon l'une quelconque des revendications 14 à 26.

28. Instructions exécutables par un processeur pour commander un processeur afin qu'il exécute le procédé selon l'une quelconque des revendications 14 à 26.

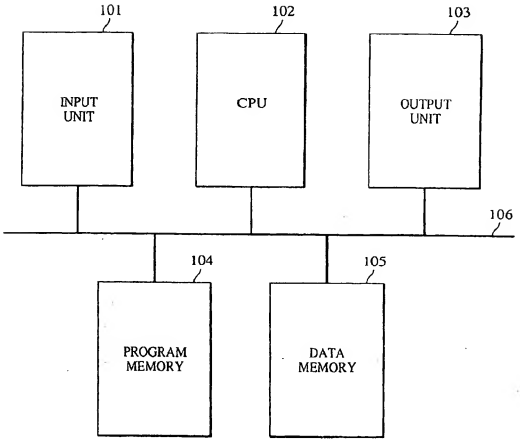

101
102
103
INPUT UNIT
CPU
OUTPUT UNIT
106
104
105
PROGRAM MEMORY
DATA MEMORY

FIG. 1

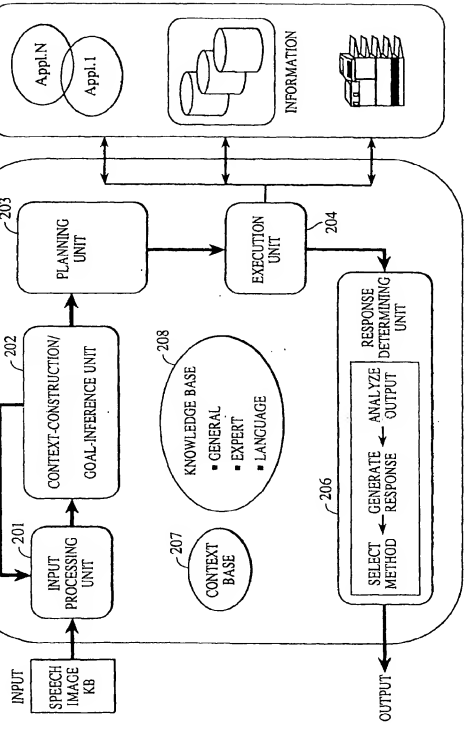

INPUT
SPEECH
IMAGE
KB
201
INPUT
PROCESSING
UNIT
202
CONTEXT-CONSTRUCTION/
GOAL-INFERENCE UNIT
203
PLANNING
UNIT
204
EXECUTION
UNIT
205
Appl.N
Appl.1
INFORMATION
206
SELECT
METHOD
GENERATE
RESPONSE
ANALYZE
OUTPUT
RESPONSE
DETERMINING
UNIT
207
CONTEXT
BASE
208
KNOWLEDGE BASE
▪ GENERAL
▪ EXPERT
▪ LANGUAGE
OUTPUT

FIG. 2

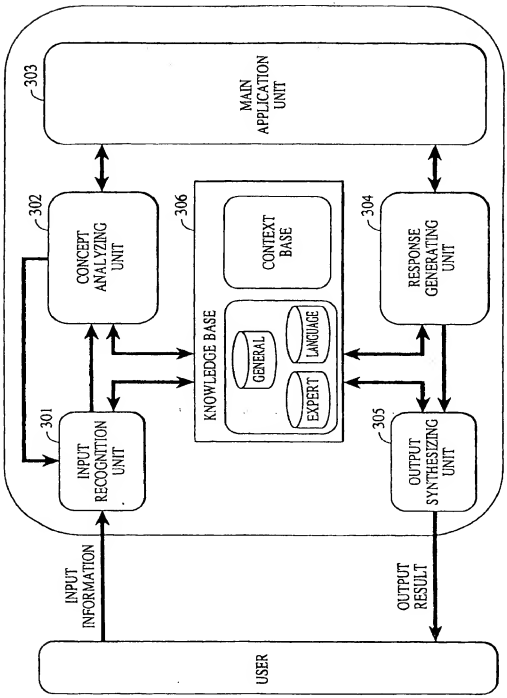

USER
INPUT INFORMATION
OUTPUT RESULT
INPUT RECOGNITION UNIT
301
CONCEPT ANALYZING UNIT
302
MAIN APPLICATION UNIT
303
RESPONSE GENERATING UNIT
304
OUTPUT SYNTHESIZING UNIT
305
KNOWLEDGE BASE
306
GENERAL
LANGUAGE
EXPERT
CONTEXT BASE

FIG. 3

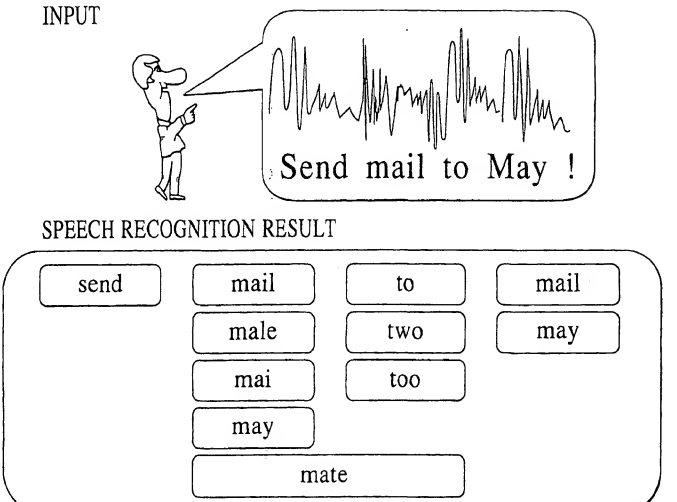

INPUT
Send mail to May !
SPEECH RECOGNITION RESULT
send
mail
male
mai
may
mate
to
two
too
mail
may

FIG. 4

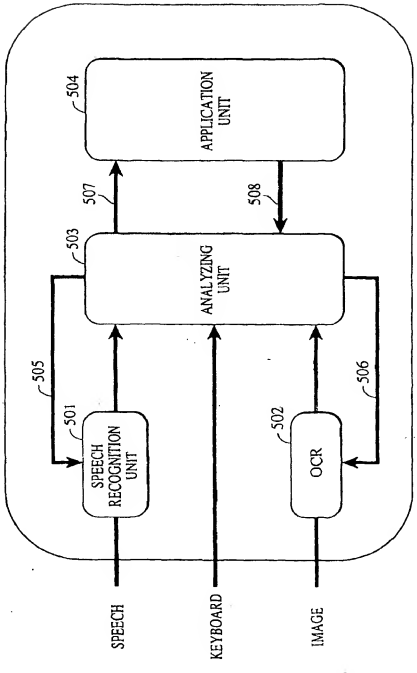

SPEECH
KEYBOARD
IMAGE
SPEECH RECOGNITION UNIT
OCR
ANALYZING UNIT
APPLICATION UNIT
501
502
503
504
505
506
507
508

FIG. 5

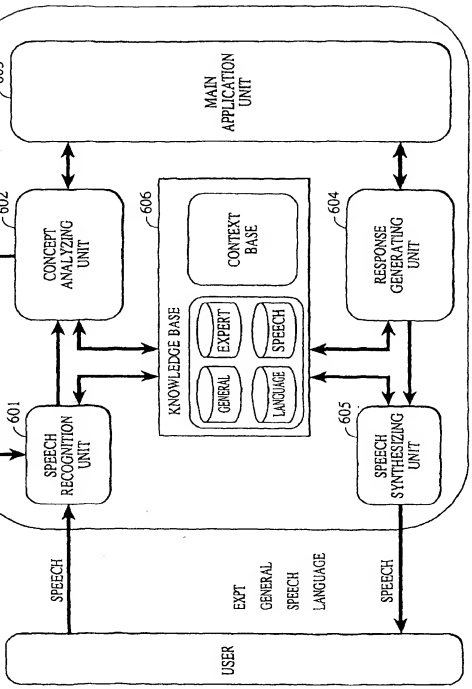

603
MAIN APPLICATION UNIT
602
CONCEPT ANALYZING UNIT
606
KNOWLEDGE BASE
CONTEXT BASE
EXPERT
SPEECH
GENERAL
LANGUAGE
604
RESPONSE GENERATING UNIT
601
SPEECH RECOGNITION UNIT
605
SPEECH SYNTHESIZING UNIT
SPEECH
EXPT
GENERAL
SPEECH
LANGUAGE
SPEECH
USER

FIG. 6

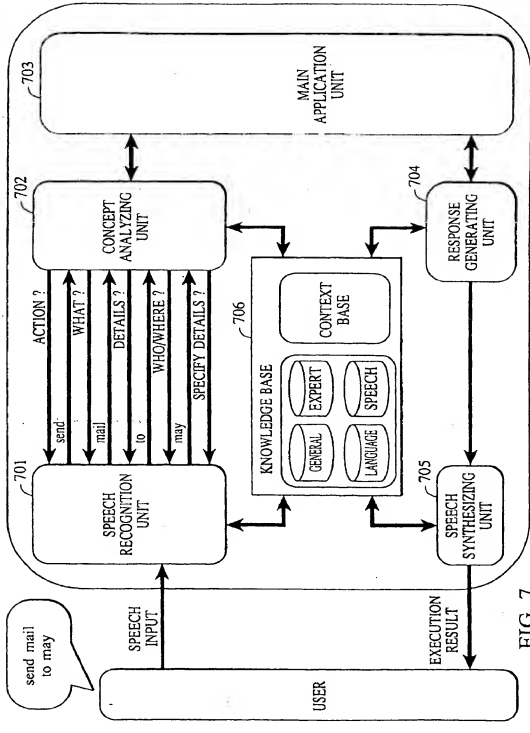

send mail to may
SPEECH INPUT
USER
EXECUTION RESULT
701
SPEECH RECOGNITION UNIT
702
CONCEPT ANALYZING UNIT
703
MAIN APPLICATION UNIT
704
RESPONSE GENERATING UNIT
705
SPEECH SYNTHESIZING UNIT
706
KNOWLEDGE BASE
GENERAL
EXPERT
SPEECH
LANGUAGE
CONTEXT BASE
ACTION ?
WHAT ?
DETAILS ?
WHO/WHERE ?
SPECIFY DETAILS ?
send
mail
to
may

FIG. 7

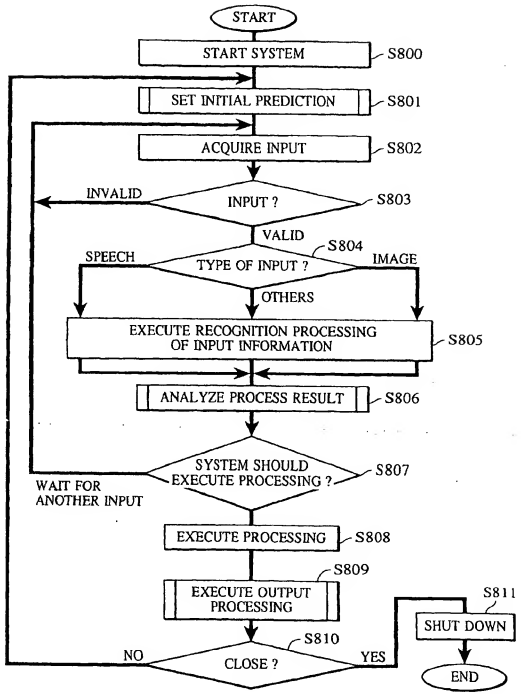

START
START SYSTEM
S800
SET INITIAL PREDICTION
S801
ACQUIRE INPUT
S802
INVALID
INPUT ?
S803
VALID
S804
SPEECH
TYPE OF INPUT ?
IMAGE
OTHERS
EXECUTE RECOGNITION PROCESSING OF INPUT INFORMATION
S805
ANALYZE PROCESS RESULT
S806
SYSTEM SHOULD EXECUTE PROCESSING ?
S807
WAIT FOR ANOTHER INPUT
EXECUTE PROCESSING
S808
S809
EXECUTE OUTPUT PROCESSING
S811
SHUT DOWN
S810
NO
CLOSE ?
YES
END

FIG. 8

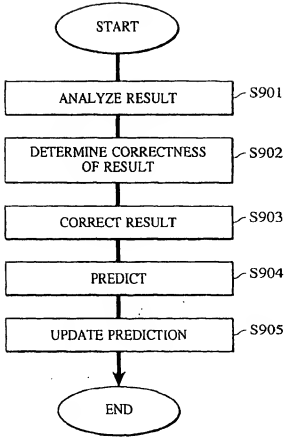

START
ANALYZE RESULT
S901
DETERMINE CORRECTNESS OF RESULT
S902
CORRECT RESULT
S903
PREDICT
S904
UPDATE PREDICTION
S905
END

FIG. 9

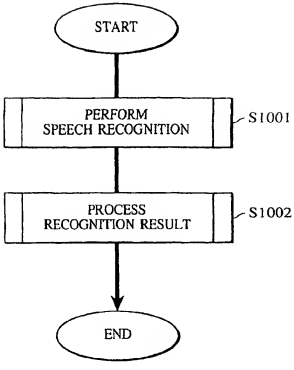

START
PERFORM SPEECH RECOGNITION
S1001
PROCESS RECOGNITION RESULT
S1002
END

FIG. 10

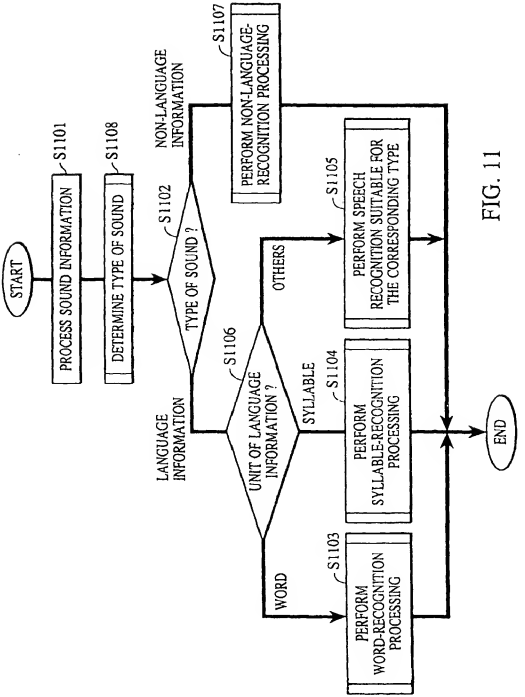

START
PROCESS SOUND INFORMATION
S1101
DETERMINE TYPE OF SOUND
S1108
TYPE OF SOUND ?
S1102
LANGUAGE INFORMATION
NON-LANGUAGE INFORMATION
S1107
PERFORM NON-LANGUAGE-RECOGNITION PROCESSING
UNIT OF LANGUAGE INFORMATION ?
S1106
WORD
SYLLABLE
OTHERS
S1103
PERFORM WORD-RECOGNITION PROCESSING
S1104
PERFORM SYLLABLE-RECOGNITION PROCESSING
S1105
PERFORM SPEECH RECOGNITION SUITABLE FOR THE CORRESPONDING TYPE
END

FIG. 11

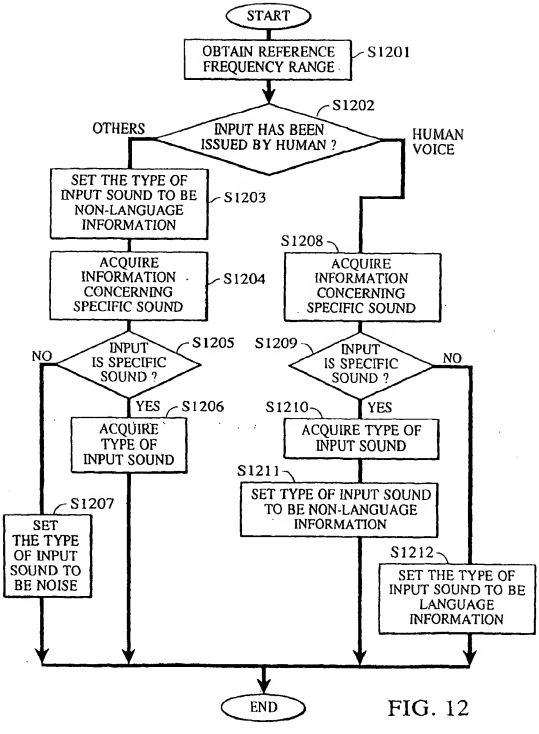

START
OBTAIN REFERENCE FREQUENCY RANGE
S1201
S1202
INPUT HAS BEEN ISSUED BY HUMAN ?
OTHERS
HUMAN VOICE
SET THE TYPE OF INPUT SOUND TO BE NON-LANGUAGE INFORMATION
S1203
ACQUIRE INFORMATION CONCERNING SPECIFIC SOUND
S1204
S1208
ACQUIRE INFORMATION CONCERNING SPECIFIC SOUND
INPUT IS SPECIFIC SOUND ?
S1205
NO
YES
S1206
ACQUIRE TYPE OF INPUT SOUND
S1207
SET THE TYPE OF INPUT SOUND TO BE NOISE
S1209
INPUT IS SPECIFIC SOUND ?
NO
YES
S1210
ACQUIRE TYPE OF INPUT SOUND
S1211
SET TYPE OF INPUT SOUND TO BE NON-LANGUAGE INFORMATION
S1212
SET THE TYPE OF INPUT SOUND TO BE LANGUAGE INFORMATION
END

FIG. 12

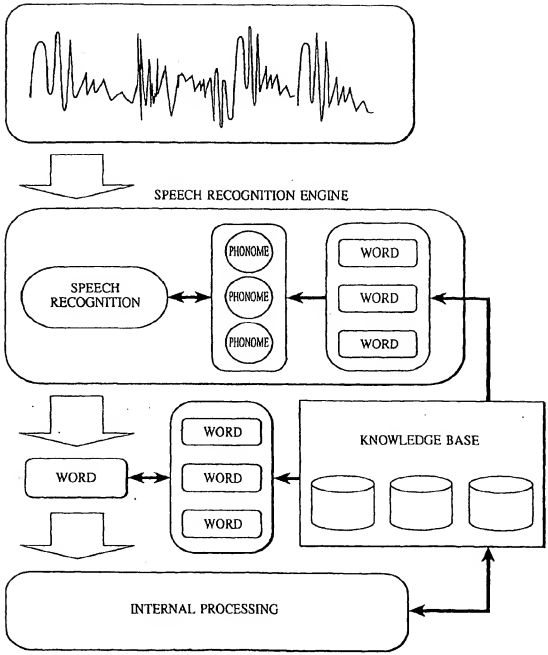

SPEECH RECOGNITION ENGINE
SPEECH RECOGNITION
PHONOME
PHONOME
PHONOME
WORD
WORD
WORD
WORD
WORD
WORD
WORD
KNOWLEDGE BASE
INTERNAL PROCESSING

FIG. 13

```
Act
  action     "#' ? @ k . S $ n #";
  access     "#' ? @ k . S $ n #";
Act2
  created    "#' k R+ i . ' e&I . t I d #";
  modified   "#' m A . d $ . f a&I d #";
Prep
  at                    "#' ? @ t #";
  by                    "#' b a&I #";
Prep2
  after                 "#' ? @ f . t $ R+ #";
  before                "#' b I . ' f O R+ #";
  till                  "#' t I l #";
  until                 "#' ? $ n . ' t I l #";
Prep3
      of                "#' ? ^ v #";
Adv
  all                   "#' ? O I #";
  new                   "#' n u #";
  previous   "#' p R+ i . v i . $ s #";
Adj
  immediately           "#' ? I . ' m i . d j $ t . l i #";
Object
  document              "#' d A . k j $ . m E n t #";
  email                 "#' ? i . m e&I l #";
  file                  "#' f a&I l #";
  mail                  "#' m e&I l #";
  list                  "#' l I s t #";
```

FIG. 14

Name
- *Person*
- *Machine*

Time
- *day*
- *month*

Number
- *cardinal number*
- *ordinal number*

Alphabet

FIG. 15

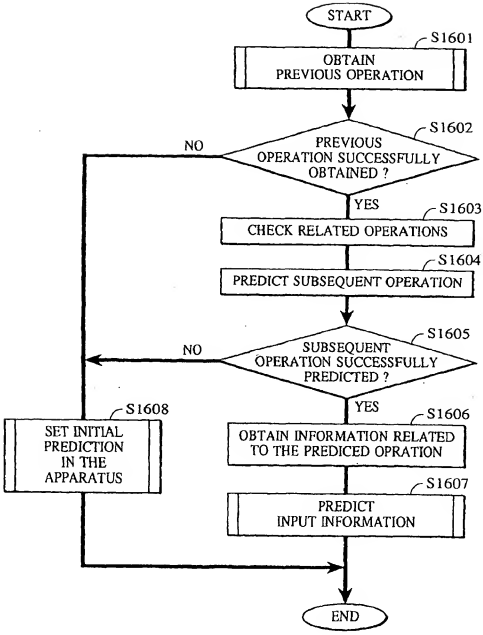

START
S1601
OBTAIN PREVIOUS OPERATION
S1602
PREVIOUS OPERATION SUCCESSFULLY OBTAINED ?
NO
YES
S1603
CHECK RELATED OPERATIONS
S1604
PREDICT SUBSEQUENT OPERATION
S1605
SUBSEQUENT OPERATION SUCCESSFULLY PREDICTED ?
NO
YES
S1606
OBTAIN INFORMATION RELATED TO THE PREDICED OPRATION
S1607
PREDICT INPUT INFORMATION
S1608
SET INITIAL PREDICTION IN THE APPARATUS
END

FIG. 16

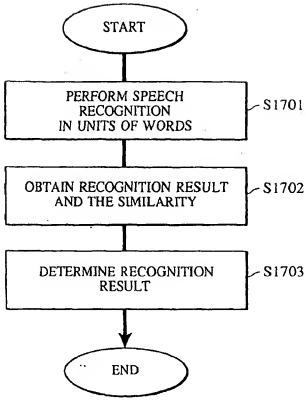

START
PERFORM SPEECH RECOGNITION IN UNITS OF WORDS
S1701
OBTAIN RECOGNITION RESULT AND THE SIMILARITY
S1702
DETERMINE RECOGNITION RESULT
S1703
END

FIG. 17

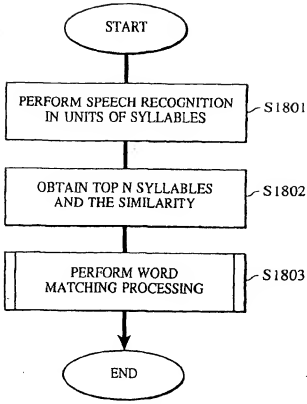

START
PERFORM SPEECH RECOGNITION IN UNITS OF SYLLABLES
S1801
OBTAIN TOP N SYLLABLES AND THE SIMILARITY
S1802
PERFORM WORD MATCHING PROCESSING
S1803
END

FIG. 18

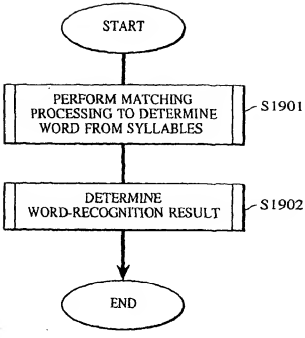

START
PERFORM MATCHING PROCESSING TO DETERMINE WORD FROM SYLLABLES
S1901
DETERMINE WORD-RECOGNITION RESULT
S1902
END

FIG. 19

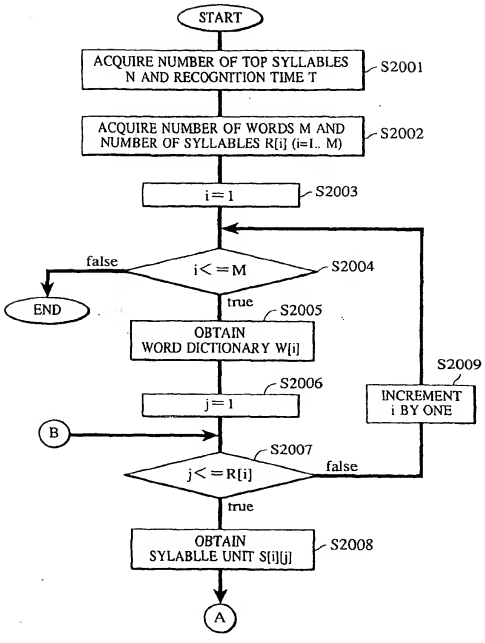

START
ACQUIRE NUMBER OF TOP SYLLABLES N AND RECOGNITION TIME T
S2001
ACQUIRE NUMBER OF WORDS M AND NUMBER OF SYLLABLES R[i] (i=1.. M)
S2002
i=1
S2003
i<=M
S2004
false
END
true
OBTAIN WORD DICTIONARY W[i]
S2005
j=1
S2006
B
j<=R[i]
S2007
false
true
OBTAIN SYLABLLE UNIT S[i][j]
S2008
A
INCREMENT i BY ONE
S2009

FIG. 20

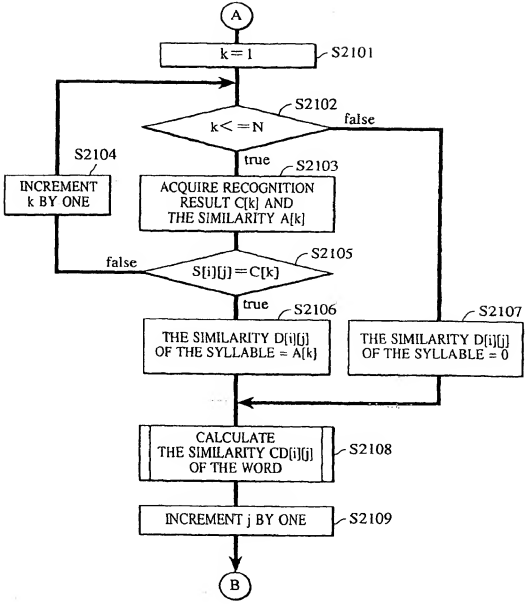

A
k=1
S2101
S2102
k<=N
false
true
S2104
INCREMENT k BY ONE
S2103
ACQUIRE RECOGNITION RESULT C[k] AND THE SIMILARITY A[k]
S2105
S[i][j]=C[k]
false
true
S2106
THE SIMILARITY D[i][j] OF THE SYLLABLE = A[k]
S2107
THE SIMILARITY D[i][j] OF THE SYLLABLE = 0
CALCULATE THE SIMILARITY CD[i][j] OF THE WORD
S2108
INCREMENT j BY ONE
S2109
B

FIG. 21

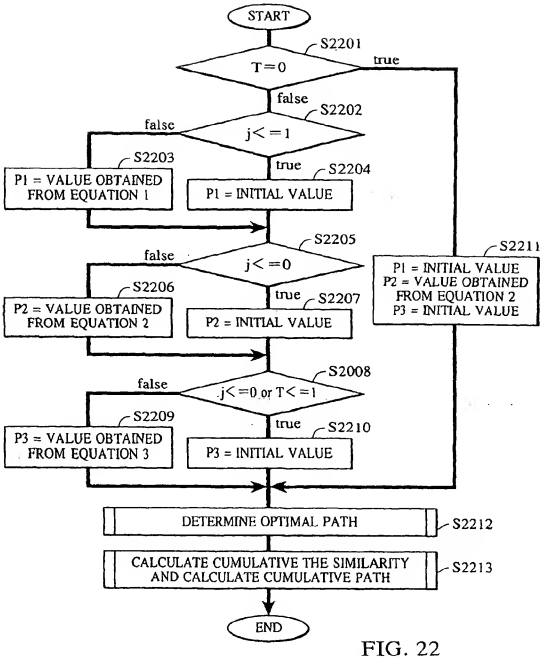

START
S2201
T=0
true
false
S2202
j<=1
false
S2203
P1 = VALUE OBTAINED FROM EQUATION 1
true
S2204
P1 = INITIAL VALUE
S2205
j<=0
false
S2206
P2 = VALUE OBTAINED FROM EQUATION 2
true
S2207
P2 = INITIAL VALUE
S2211
P1 = INITIAL VALUE
P2 = VALUE OBTAINED FROM EQUATION 2
P3 = INITIAL VALUE
S2008
j<=0 or T<=1
false
S2209
P3 = VALUE OBTAINED FROM EQUATION 3
true
S2210
P3 = INITIAL VALUE
DETERMINE OPTIMAL PATH
S2212
CALCULATE CUMULATIVE THE SIMILARITY AND CALCULATE CUMULATIVE PATH
S2213
END

FIG. 22

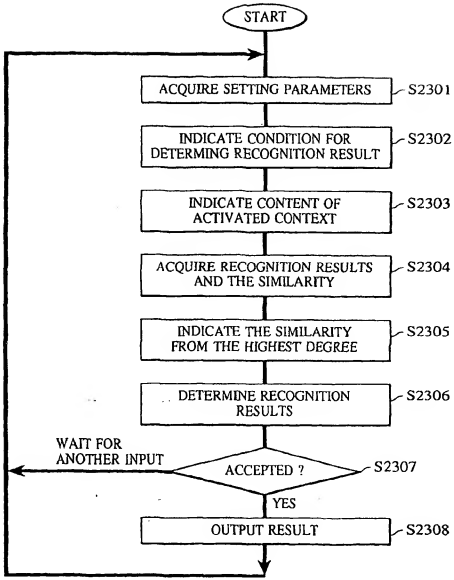

START
ACQUIRE SETTING PARAMETERS
S2301
INDICATE CONDITION FOR DETERMING RECOGNITION RESULT
S2302
INDICATE CONTENT OF ACTIVATED CONTEXT
S2303
ACQUIRE RECOGNITION RESULTS AND THE SIMILARITY
S2304
INDICATE THE SIMILARITY FROM THE HIGHEST DEGREE
S2305
DETERMINE RECOGNITION RESULTS
S2306
WAIT FOR ANOTHER INPUT
ACCEPTED ?
S2307
YES
OUTPUT RESULT
S2308

FIG. 23

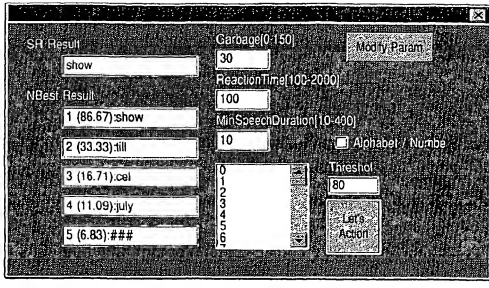

SR Result
show
NBest Result
1 (86.67):show
2 (33.33):till
3 (16.71):cel
4 (11.09):july
5 (6.83):###
Garbage[0:150]
30
ReactionTime[100:2000]
100
MinSpeechDuration[10:400]
10
Alphabet / Numbe
Threshol
80
Modify Param
Let's Action
0
1
2
3
4
5
6

FIG. 24

SYLLABLE DICTIONARY

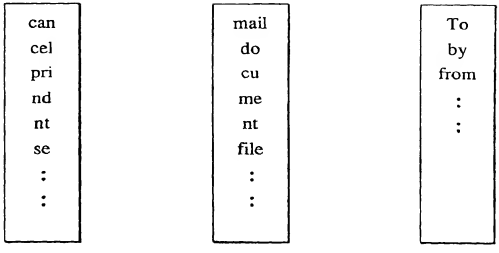


FIG. 25

WORD DICTIONARY

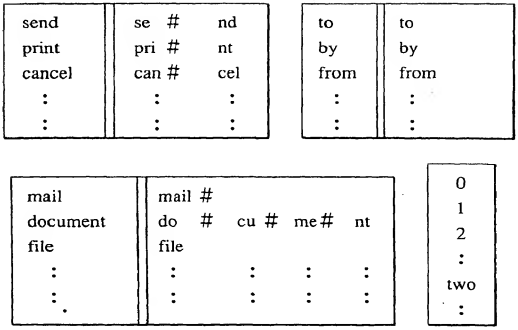


FIG. 26

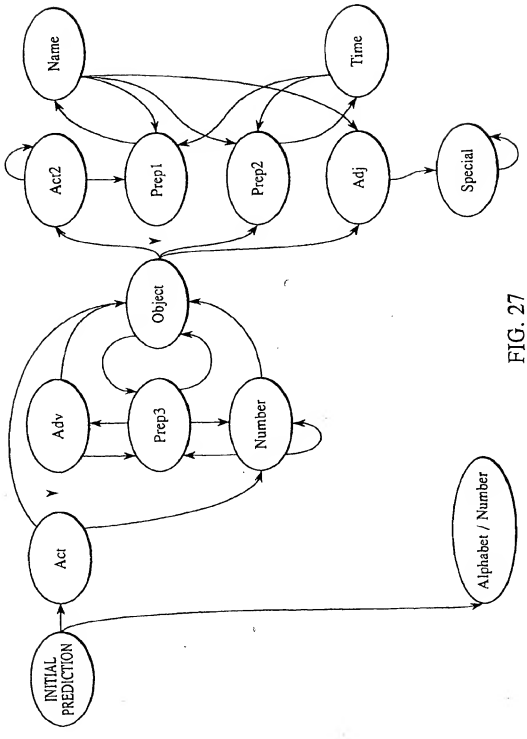

INITIAL PREDICTION
Act
Adv
Prep3
Number
Object
Act2
Prep1
Prep2
Adj
Name
Time
Special
Alphabet / Number

FIG. 27

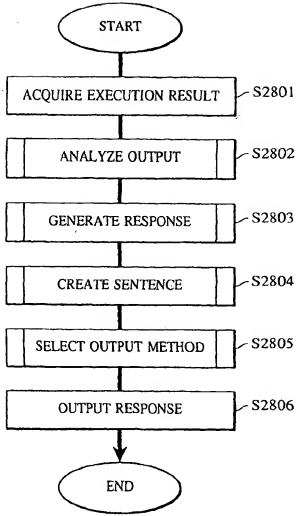

START
ACQUIRE EXECUTION RESULT
S2801
ANALYZE OUTPUT
S2802
GENERATE RESPONSE
S2803
CREATE SENTENCE
S2804
SELECT OUTPUT METHOD
S2805
OUTPUT RESPONSE
S2806
END

FIG. 28

LANGUAGE

```
Send : Verb wform Obj Iobj *Instrument
Print : Verb wform Obj Iobj
Mail : noun countable
...
```

FIG. 29

CONCEPT

```
Send : PTRANS
Print : PACT
Mail : DOCUMENT
...
```

FIG. 30

RULE

Send : to May Print : at Printer1 Mail : From John ...

FIG. 31

WORD DICTIONARY

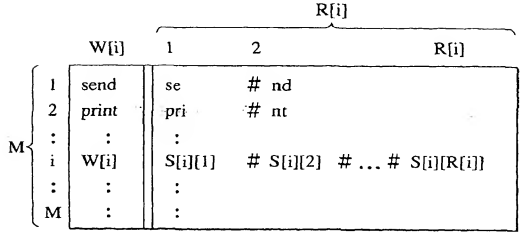
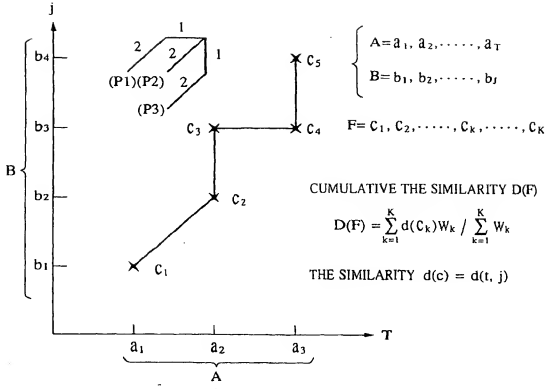


FIG. 32

SYLLABLE RECOGNITION RESULT

		C[k]	A[k]
N	1	nd	60.4
	2	pri	53.7
	:	:	:
	k	C[k]	A[k]
	:	:	:
	N	:	:

FIG. 33



CUMULATIVE THE SIMILARITY (STATE VOLUMN) g(t, j)

$$g(t,j)=\max\begin{bmatrix} g(t-2,\ j-1)+2d(t-1,\ j)+d(t,\ j)\Lambda\ (1) \\ g(t-1,\ j-1)+2d(t,\ j)\Lambda\ (2) \\ g(t-1,\ j-2)+2d(t,\ j-1)+d(t,\ j)\Lambda\ (3) \end{bmatrix}$$

LENGTH OF OPTIMAL PATH c(t, j)

$$c(t,j)=\max\begin{bmatrix} c(t-2,\ j-1)+3\Lambda \text{ if } (1) \\ c(t-1,\ j-1)+2\Lambda \text{ if } (2) \\ c(t-1,\ j-2)+3\Lambda \text{ if } (3) \end{bmatrix}$$

FIG. 34

SYLLABLE RECOGNITION RESULT

T=0

N=5	1	nd	61.0
	2	se	60.4
	3	pri	53.7
	4	com	29.1
	5	nt	20.6

T=1

N=1	nd	100.0

FIG. 35

WORD DICTIONARY

			R[i]			
		W[i]	1	2	3	4
M=3	1	send	se	# nd		
	2	print	pri	# nt		
	3	communicate	com	# mu	# ni	# cate

FIG. 36

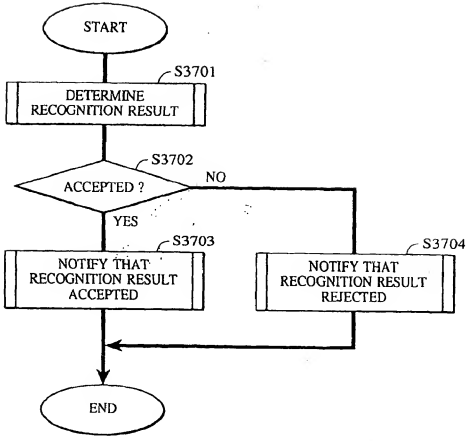

START
S3701
DETERMINE RECOGNITION RESULT
S3702
ACCEPTED ?
NO
YES
S3703
NOTIFY THAT RECOGNITION RESULT ACCEPTED
S3704
NOTIFY THAT RECOGNITION RESULT REJECTED
END

FIG. 37

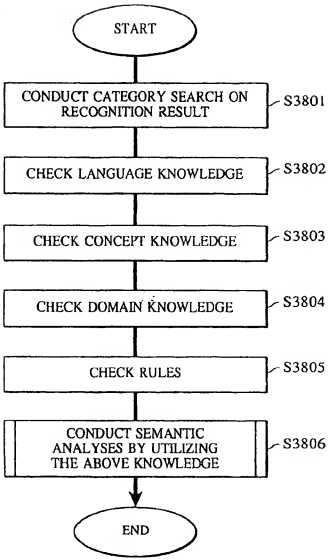

START
CONDUCT CATEGORY SEARCH ON RECOGNITION RESULT
S3801
CHECK LANGUAGE KNOWLEDGE
S3802
CHECK CONCEPT KNOWLEDGE
S3803
CHECK DOMAIN KNOWLEDGE
S3804
CHECK RULES
S3805
CONDUCT SEMANTIC ANALYSES BY UTILIZING THE ABOVE KNOWLEDGE
S3806
END

FIG. 38

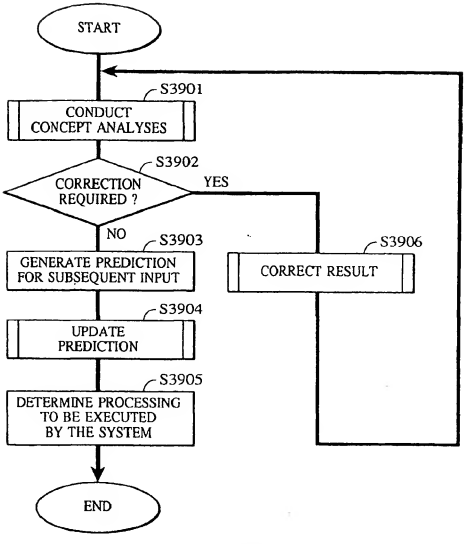

START
S3901
CONDUCT CONCEPT ANALYSES
S3902
CORRECTION REQUIRED ?
YES
NO
S3903
GENERATE PREDICTION FOR SUBSEQUENT INPUT
S3904
UPDATE PREDICTION
S3905
DETERMINE PROCESSING TO BE EXECUTED BY THE SYSTEM
END
S3906
CORRECT RESULT

FIG. 39

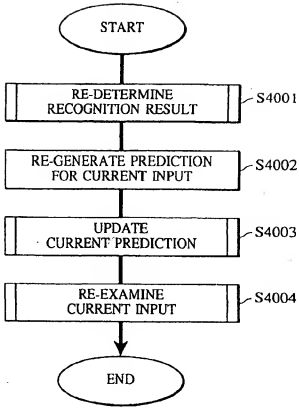

START
RE-DETERMINE RECOGNITION RESULT
S4001
RE-GENERATE PREDICTION FOR CURRENT INPUT
S4002
UPDATE CURRENT PREDICTION
S4003
RE-EXAMINE CURRENT INPUT
S4004
END

FIG. 40

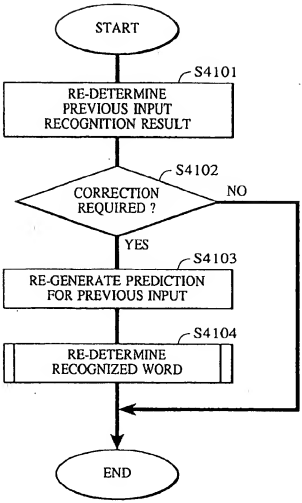

START
S4101
RE-DETERMINE PREVIOUS INPUT RECOGNITION RESULT
S4102
CORRECTION REQUIRED ?
NO
YES
S4103
RE-GENERATE PREDICTION FOR PREVIOUS INPUT
S4104
RE-DETERMINE RECOGNIZED WORD
END

FIG. 41

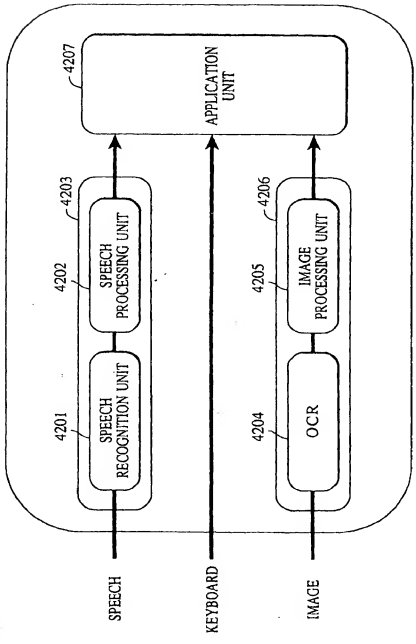

4207
APPLICATION UNIT
4203
4202
SPEECH PROCESSING UNIT
4206
4205
IMAGE PROCESSING UNIT
4201
SPEECH RECOGNITION UNIT
4204
OCR
SPEECH
KEYBOARD
IMAGE

FIG. 42